

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



**Cincuenta
Aniversario**

UAM Universidad Autónoma
de Madrid

Grado en Ingeniería Informática

TRABAJO FIN DE GRADO

**Réplica y Agregación de Resultados de un Experimento
Verdadero sobre el Impacto de los Mecanismos de
Usabilidad de Preferencias, Retroalimentación de Progreso
y Abortar Operación en un Entorno Web**

Adriana Iglesias Pedrejón
Tutora: Silvia Teresita Acuña Castillo

JUNIO 2018

**Réplica y Agregación de Resultados de un Experimento
Verdadero sobre el Impacto de los Mecanismos de
Usabilidad de Preferencias, Retroalimentación de Progreso
y Abortar Operación en un Entorno Web**

**AUTORA: Adriana Iglesias Pedrejón
TUTORA: Silvia Teresita Acuña Castillo**

**Grupo de Investigación de Herramientas Interactivas Avanzadas (GHIA)
Departamento de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio de 2018**

Resumen

La usabilidad es un atributo de calidad clave del sistema software. Sin embargo, no se encuentran en la literatura trabajos que agreguen los resultados de réplicas de experimentos verdaderos y obtengan evidencias empíricas consolidadas sobre el impacto de la presencia-ausencia de mecanismos de usabilidad en aplicaciones web.

El objetivo de este trabajo es realizar una réplica de un experimento verdadero que analiza cómo afectan los mecanismos de usabilidad: Abortar Operación, Retroalimentación de Progreso y Preferencias, a la eficiencia (tiempo y número de clics para realizar la tarea), la eficacia (porcentaje de completitud de la tarea) y la satisfacción de los sujetos al enfrentarse a las tareas que conforman el test de usabilidad en un entorno web.

Además, en esta investigación se agregan los resultados del experimento original llevado a cabo en 2016, su primera réplica en 2017 y la segunda réplica realizada en este trabajo, y se aplica el modelo de regresión lineal para evaluar estos tres conjuntos de 267 datos a nivel de familia por cada mecanismo de usabilidad.

Una vez realizado el experimento se plantea un estudio estadístico que comienza analizando los estadísticos descriptivos y los correspondientes diagramas de violín para las distintas combinaciones de factores. A continuación, se realiza la prueba U-Mann Whitney donde se calcula el p-valor, el tamaño de efecto Cliff's delta, la evaluación de la significación estadística y del tamaño de efecto y la probabilidad de superioridad de un grupo sobre otro para cada variable respuesta y mecanismo.

En el caso de la variable respuesta *eficiencia*, aunque la existencia del mecanismo de usabilidad nunca perjudica la eficiencia de los usuarios, no siempre se observa una mejora determinante. En el caso del mecanismo Abortar Operación, el tamaño de efecto es pequeño por lo que no parece ser relevante y, en consecuencia, la mejora en la eficiencia del usuario es mínima. En el caso del mecanismo Retroalimentación de Progreso, aunque el tamaño de efecto es medio para el número de clics, la diferencia entre la ausencia y presencia del mecanismo tampoco es concluyente. Por último, para el mecanismo Preferencias, la diferencia entre la presencia o ausencia del mecanismo es irrelevante.

En el caso de la variable respuesta *eficacia* para los mecanismos Abortar Operación y Preferencias la diferencia entre la presencia y ausencia de cada mecanismo es grande, por lo que la mejora es concluyente. En el mecanismo Retroalimentación de Progreso no se aprecia una mejora significativa.

Para la variable respuesta *satisfacción* tanto en el mecanismo Abortar Operación como Preferencias existe diferencia significativa entre la ausencia y presencia del mecanismo, mientras que en el caso de Retroalimentación de Progreso la mejora es prácticamente inexistente.

Por último, los resultados del meta-análisis nos permiten concluir que la presencia o ausencia de los mecanismos Abortar Operación y Preferencias parecen afectar en gran medida a la mejora de la usabilidad del sistema tanto en la eficiencia, eficacia como en la

satisfacción del usuario, mientras que el mecanismo Retroalimentación de Progreso parece afectar en menor medida a la usabilidad del sistema.

Palabras clave

Usabilidad, Abortar Operación, Retroalimentación de Progreso, Preferencias, Eficiencia, Eficacia, Satisfacción, Agregación.

Abstract

Usability is a quality attribute of the software system. However, there are no works in the literature that aggregate the results of replicas of true experiments and obtain consolidated empirical evidence on the impact on the presence-absence of usability mechanisms in web applications.

The objective of this work is to make a replica of a true experiment that analyzes how usability mechanisms affect: Abort Operation, Progress Feedback and Preferences, efficiency (time and number of clicks to perform the task), efficiency (percentage of completeness of the task) and the satisfaction of the subjects when facing the tasks that make up the usability test in a web environment.

In addition, this research adds the results of the original experiment carried out in 2016, its first replication in 2017 and the second replication performed in this work, and the linear regression model is applied to evaluate these three sets of these three sets of 267 data at the family level for each usability mechanism.

Once the experiment is carried out, a statistical study is proposed, which begins by analyzing the descriptive statistics and the corresponding violin diagrams for the different combinations of factors. Next, the U-Mann Whitney test is performed, where the p-value, the Cliff's delta effect size, the evaluation of the statistical significance and the effect size and the probability of superiority of one group over another for each response variable and mechanism are calculated.

In the case of the efficiency response variable, although the existence of the usability mechanism never harms the efficiency of the users, a determining improvement is not always observed. In the case of the Abort Operation mechanism, the effect size is small, so it does not seem to be relevant and, consequently, the improvement in user efficiency is minimal. In the case of the Progress Feedback mechanism, although the effect size is average for the number of clicks, the difference between the absence and presence of the mechanism is not conclusive either. Finally, for the Preferences mechanism, the difference between the presence and absence of the mechanism is irrelevant.

In the case of the effectiveness response variable for the Abort Operation and Preferences mechanisms, the difference between the presence and absence of each mechanism is large, so the improvement is conclusive. There is no significant improvement in the Progress Feedback mechanism.

For the satisfaction response variable in both the Abort Operation and Preferences mechanism, there is a significant difference between the absence and presence of the mechanism, while in the case of Progress Feedback, the improvement is practically non-existent.

Finally, the results of the meta-analysis allow us to conclude that the presence or absence of the Abort Operation and Preferences mechanisms seem to greatly affect the improvement of the usability of the system in terms of efficiency, effectiveness and user satisfaction. While the Progress Feedback mechanism seems to affect to a lesser extent the usability of the system.

Keywords

*Usability, Abort Operation, Progress Feedback, Preferences, Efficiency, Effectiveness
Satisfaction, Aggregation.*

Agradecimientos

Agradezco a todas las personas que me han acompañado durante esta etapa tan importante en mi vida. En primer lugar, a mi tutora Silvia, que ha estado siempre disponible para ayudarme y enseñarme durante este proceso.

Agradecer también a Marcelo y a Adrián, que aunque no me conocían, han estado siempre dispuestos a echarme una mano en lo que he necesitado.

También quiero acordarme de los alumnos de la Universidad Autónoma de Asunción de Paraguay que colaboraron en la realización de este trabajo de forma generosa.

Gracias a mi familia y amigos, por su cariño y apoyo incondicional.

ÍNDICE DE CONTENIDOS

1. INTRODUCCIÓN	1
1.1 VISIÓN GENERAL DE LA INVESTIGACIÓN	1
1.2 OBJETIVOS DEL TRABAJO.....	1
1.3 MÉTODO DE LA INVESTIGACIÓN	2
1.4 ESTRUCTURA DEL TRABAJO.....	2
2. ESTADO DE LA CUESTIÓN	5
2.1 EVALUACIÓN DE LA USABILIDAD	5
2.2 CARACTERÍSTICAS DE USABILIDAD FUNCIONALES.....	6
2.3 TRABAJOS RELACIONADOS	6
3. EXPERIMENTO VERDADERO	9
3.1 OBJETIVO, INVESTIGACIÓN E HIPÓTESIS	9
3.2 FACTORES Y VARIABLES RESPUESTA	10
3.3 CONTEXTO Y SUJETOS EXPERIMENTALES	10
3.4 DISEÑO EXPERIMENTAL	11
3.5 HERRAMIENTAS Y TAREAS	13
3.6 OPERACIÓN	15
3.7 AMENAZAS A LA VALIDEZ	15
3.7.1 Validez Interna.....	15
3.7.2 Validez Externa	16
4. ENFOQUE DE ANÁLISIS.....	17
4.1 ANÁLISIS DE DATOS.....	17
4.1.1 RQ1: Abortar Operación (MU-ABR).....	17
4.1.2 RQ2: Retroalimentación de Progreso (MU-PFB).....	20
4.1.3 RQ3: Preferencias (MU-PRF).....	23
4.2 DISCUSIÓN DE RESULTADOS	26
4.2.1 RQ1: Abortar Operación (MU-ABR).....	26
4.2.2 RQ2: Retroalimentación de Progreso (MU-PFB).....	27
4.2.3 RQ3: Preferencias (MU-PRF).....	28
5. AGREGACIÓN DE RESULTADOS.....	29
5.1 ENFOQUE DE ANÁLISIS.....	29
5.2 ANÁLISIS DE ABORTAR OPERACIÓN	30
5.3 ANÁLISIS DE RETROALIMENTACIÓN DE PROGRESO	31
5.4 ANÁLISIS DE PREFERENCIAS	32
6. CONCLUSIONES Y TRABAJO FUTURO	35
6.1 CONCLUSIÓN	35
6.2 TRABAJO FUTURO.....	36
REFERENCIAS.....	37
GLOSARIO	41
ANEXOS.....	43

A.	CUESTIONARIO DE FAMILIARIDAD	43
B.	DISEÑO DE TAREAS Y CUESTIONARIOS DE SATISFACCIÓN POST-TAREA	45
C.	DIAGRAMAS DE CAJAS.....	47
	C.1 DIAGRAMA DE CAJAS DEL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN	47
	C.2 DIAGRAMA DE CAJAS DEL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	50
	C.3 DIAGRAMA DE CAJAS DEL MECANISMO DE USABILIDAD PREFERENCIAS	52

ÍNDICE DE FIGURAS

FIGURA 1: FASES DE LA INVESTIGACIÓN	2
FIGURA 2: DIAGRAMA DE VIOLÍN PARA EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	19
FIGURA 3: DIAGRAMA DE VIOLÍN PARA EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO	22
FIGURA 4: DIAGRAMA DE VIOLÍN PARA EL MECANISMO DE USABILIDAD PREFERENCIAS	25
FIGURA 5 : GRÁFICO DE PERFIL PARA ABORTAR OPERACIÓN.....	30
FIGURA 6: GRÁFICO DE PERFIL PARA RETROALIMENTACIÓN DE PROGRESO	31
FIGURA 7: GRÁFICO DE PERFIL PARA PREFERENCIAS.....	32
FIGURA 8: CUESTIONARIO DE FAMILIARIDAD DE LA APLICACIÓN QUICKSTORE	43
FIGURA 9: TAREA PROPUESTA Y CUESTIONARIO POST-TAREA PARA EL MU-ABR.....	45
FIGURA 10: TAREA PROPUESTA Y CUESTIONARIO POST-TAREA PARA EL MU-PFB	46
FIGURA 11: TAREA, SUBTAREA Y CUESTIONARIO POST-TAREA PARA EL MU-PRF	46
FIGURA 12: DIAGRAMA DE CAJAS PARA LA VARIABLE N° DE CLICS EN EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	47
FIGURA 13: DIAGRAMA DE CAJAS PARA LA VARIABLE TIEMPO EN EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	48
FIGURA 14: DIAGRAMA DE CAJAS PARA LA VARIABLE PORCENTAJE DE TAREA EN EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	49
FIGURA 15: DIAGRAMA DE CAJAS PARA LA VARIABLE SATISFACCIÓN EN EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	49
FIGURA 16: DIAGRAMA DE CAJAS PARA LA VARIABLE N° DE CLICS EN EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	50
FIGURA 17: DIAGRAMA DE CAJAS PARA LA VARIABLE TIEMPO EN EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	51
FIGURA 18: DIAGRAMA DE CAJAS PARA LA VARIABLE PORCENTAJE DE TAREA EN EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	51
FIGURA 19: DIAGRAMA DE CAJAS PARA LA VARIABLE SATISFACCIÓN EN EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	52
FIGURA 20: DIAGRAMA DE CAJAS PARA LA VARIABLE N° DE CLICS EN EL MECANISMO DE USABILIDAD PREFERENCIAS	53
FIGURA 21: DIAGRAMA DE CAJAS PARA LA VARIABLE TIEMPO EN EL MECANISMO DE USABILIDAD PREFERENCIAS	53
FIGURA 22: DIAGRAMA DE CAJAS PARA LA VARIABLE PORCENTAJE DE TAREA EN EL MECANISMO DE USABILIDAD PREFERENCIAS	54
FIGURA 23: DIAGRAMA DE CAJAS PARA LA VARIABLE SATISFACCIÓN EN EL MECANISMO DE USABILIDAD PREFERENCIAS	55

ÍNDICE DE TABLAS

TABLA 1: MATRIZ DE TRATAMIENTO	12
TABLA 2: MATRIZ DE ORDEN DE EXPOSICIÓN	12
TABLA 3: MATRIZ DE ASIGNACIÓN DE GRUPOS	13
TABLA 4: ESTADÍSTICOS DESCRIPTIVOS PARA LA EFICIENCIA EN EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	18
TABLA 5: ESTADÍSTICOS DESCRIPTIVOS PARA LA EFICIENCIA EN EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	18
TABLA 6: ESTADÍSTICOS DESCRIPTIVOS PARA LA SATISFACCIÓN EN EL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN.....	18
TABLA 7: DATOS ESTADÍSTICOS DEL MECANISMO DE USABILIDAD ABORTAR OPERACIÓN	20
TABLA 8: ESTADÍSTICOS DESCRIPTIVOS PARA LA EFICIENCIA EN EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	21
TABLA 9: ESTADÍSTICOS DESCRIPTIVOS PARA LA EFICACIA EN EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	21
TABLA 10: ESTADÍSTICOS DESCRIPTIVOS PARA LA SATISFACCIÓN EN EL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO.....	21
TABLA 11: DATOS ESTADÍSTICOS DEL MECANISMO DE USABILIDAD RETROALIMENTACIÓN DE PROGRESO	23
TABLA 12: ESTADÍSTICOS DESCRIPTIVOS PARA LA EFICIENCIA EN EL MECANISMO DE USABILIDAD PREFERENCIAS...	24
TABLA 13: ESTADÍSTICOS DESCRIPTIVOS PARA LA EFICACIA EN EL MECANISMO DE USABILIDAD PREFERENCIAS	24
TABLA 14: ESTADÍSTICOS DESCRIPTIVOS PARA LA SATISFACCIÓN EN EL MECANISMO DE USABILIDAD PREFERENCIAS	24
TABLA 15: DATOS ESTADÍSTICOS DEL MECANISMO DE USABILIDAD PREFERENCIAS	25
TABLA 16: RESUMEN DE LOS RESULTADOS EXPERIMENTALES	26
TABLA 17: COEFICIENTES DE REGRESIÓN LINEAL PARA ABORTAR OPERACIÓN.....	31
TABLA 18: COEFICIENTES DE REGRESIÓN LINEAL PARA RETROALIMENTACIÓN DE PROGRESO.....	32
TABLA 19: COEFICIENTES DE REGRESIÓN LINEAL PARA PREFERENCIAS	33

1. Introducción

1.1 Visión General de la Investigación

La usabilidad es un atributo de calidad del sistema software, que según el estándar internacional [ISO 9241-11, 1998] se refiere al grado en que un producto puede ser usado por usuarios específicos para conseguir metas particulares con eficacia, eficiencia y satisfacción dado un contexto determinado de uso. Según [Juristo et al., 2007a; Winter & Rönkkö, 2010; Battey, 2001; Donahue, 2001; Black, 2002] tener en cuenta este atributo de calidad proporciona varios beneficios añadidos: mejoras en la relación coste-beneficio con su consiguiente aumento en los ingresos de proyectos software, mejoras en la productividad del equipo de trabajo, facilidad en el mantenimiento de los productos, etc.

En esta investigación se considerarán tres mecanismos de usabilidad en un entorno web con grandes características interactivas. Concretamente, los mecanismos de Abortar Operación, Retroalimentación de Progreso y Preferencias. Debido a las características interactivas del sistema software que se utilizará, dichos mecanismos de usabilidad pueden llegar a ser críticos por su gran impacto en el diseño del software tal y como se destaca en [Juristo et al., 2007b].

Si bien ya se han realizado distintos estudios analizando la influencia de aplicar estos mecanismos en la usabilidad de sistemas software [Bouza, 2016; Ferreira, 2017], no se encuentran en la literatura trabajos que agreguen los resultados de réplicas de experimentos verdaderos y obtengan evidencias empíricas consolidadas sobre el impacto de la presencia-ausencia de mecanismos de usabilidad en aplicaciones web.

La réplica de estudios empíricos es fundamental para reforzar la potencia de los resultados, corroborarlos y llegar a conclusiones más sólidas [Lumley et al., 2002]. El meta-análisis es un conjunto de herramientas estadísticas utilizadas para sintetizar los datos de una colección de estudios [Borenstein et al., 2011]. Mediante la agregación de resultados, obtenemos una muestra más grande donde dejamos de ver a los experimentos de manera aislada y comenzamos a tener evidencias abaladas por más de cien sujetos, lo cual mejora en gran medida la calidad de la conclusión. No solo porque a la vista de quien analiza los resultados, las conclusiones se apoyan en un mejor nivel de evidencia, sino también porque los métodos estadísticos se comportan más eficazmente cuando los tamaños de las muestras son mayores [Amatriain et al., 2014].

1.2 Objetivos del Trabajo

El objetivo de este trabajo es realizar una réplica de un experimento verdadero que analiza cómo afectan los mecanismos de usabilidad: Abortar Operación, Retroalimentación de Progreso y Preferencias, en un entorno web. En concreto, se utilizará la aplicación *QuickStore* (<http://webadm.senado.gov.py/tesisweb/>) que es una herramienta de compra online de distintos tipos de artículos.

En este estudio empírico se evaluará la eficiencia, la eficacia y la satisfacción de los sujetos al enfrentarse a las tareas que conforman el test de usabilidad. En el Anexo B se pueden encontrar las distintas tareas y cuestionarios de los que se compone el experimento.

La obtención de tales evidencias es fundamental, ya que incorporar estos mecanismos en un sistema software requiere de un considerable esfuerzo de diseño materializado en la creación de componentes específicos (por ejemplo, clases, métodos, etc.) que han de incorporarse en los modelos de desarrollo funcionales. Por lo tanto, la obtención de estas evidencias justificará o no este esfuerzo adicional requerido.

El experimento original ha sido llevado a cabo en [Bouza, 2016] y el detalle de su primera réplica se encuentra en [Ferreira, 2017]. En este trabajo se realizará una segunda réplica, obteniendo así tres conjuntos de resultados del estudio experimental. Una vez realizada la réplica, se aplicarán técnicas de meta-análisis para comprobar la consistencia entre los resultados del experimento original y las dos réplicas del mismo, ayudando además a generalizar la evidencia del experimento original.

1.3 Método de la Investigación

De acuerdo a los objetivos planteados en esta investigación, se llevará a cabo un método que consiste en cuatro fases (ver Figura 1). En la primera fase, se comprenderá el diseño del experimento verdadero para comprobar el impacto de los mecanismos de usabilidad seleccionados: Abortar Operación, Retroalimentación de Progreso y Preferencias, en la eficiencia, la eficacia y la satisfacción de los usuarios utilizando la aplicación web *QuickStore*, diseñada para este experimento.

En la segunda fase en este trabajo, se llevará a cabo el experimento verdadero con 99 sujetos, se realizará un análisis de los datos obtenidos en el mismo y se discutirán los resultados. A continuación en la tercera fase, procederemos a la agregación de datos y su análisis quedándonos por tanto el experimento original [Bouza, 2016], la primera réplica ya realizada [Ferreira, 2017] y esta segunda réplica del presente trabajo. Por último en la cuarta fase realizaremos una discusión del conjunto de resultados.

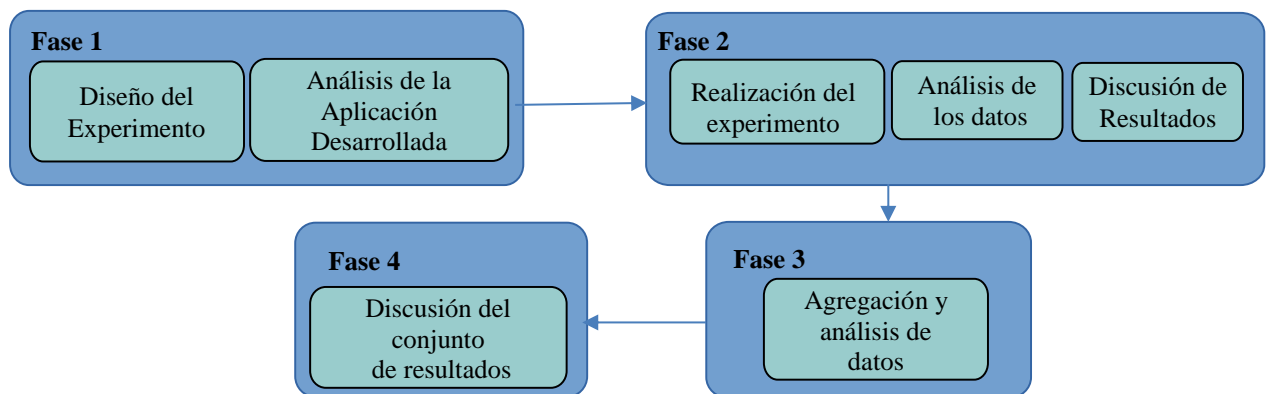


Figura 1: Fases de la Investigación

1.4 Estructura del Trabajo

Este trabajo se divide en seis capítulos, los apartados de Referencias y Glosario y cuenta además con tres Anexos. A continuación, se describirá el contenido de cada uno de ellos.

En el primer capítulo se presenta una visión general de la investigación, se plantea el problema a investigar y los objetivos del trabajo. Además, se define el método utilizado en la investigación para llegar a la solución.

En el segundo capítulo se aborda el estado de la cuestión. Se comienza analizando la evaluación de la usabilidad, dónde se menciona la importancia de la usabilidad hoy en día, los diferentes métodos que existen para estudiarla y los factores que influyen en el estudio de la misma. A continuación, se describe la clasificación de las características de usabilidad funcionales utilizadas en este trabajo y por último se discuten los trabajos relacionados que nos servirán de referencia a lo largo de esta investigación.

En el tercer capítulo se detalla principalmente el experimento llevado a cabo. Se presentan en primer lugar los objetivos y las hipótesis del mismo. A continuación, se especifica el factor principal y las variables respuesta. Asimismo, se describe el contexto y los sujetos experimentales que realizarán el experimento y posteriormente el diseño llevado a cabo junto con las herramientas y tareas que posee. Por último, se tratan las amenazas a la validez del experimento, tanto la validez interna como externa.

El cuarto capítulo muestra el análisis de datos realizado y los resultados obtenidos en la evaluación así como también la discusión de los mismos.

En el quinto capítulo se agregan los datos del experimento realizado en este trabajo, a los del experimento original y a la primera réplica. A continuación, se realiza un análisis de los datos y la discusión de los resultados obtenidos.

En el sexto capítulo se presentan las conclusiones y el trabajo futuro.

Seguidamente se presenta un apartado con las referencias utilizadas durante todo el proceso de investigación y un glosario con definiciones estadísticas aclaratorias.

En el Anexo A se presenta el cuestionario de familiaridad utilizado en el experimento verdadero final.

En el Anexo B se muestra el diseño de las tareas propuestas en el experimento, así como los cuestionarios de satisfacción post-tarea.

En el Anexo C se comentan los diagramas de caja del experimento verdadero que complementan a los resultados obtenidos en el cuarto capítulo.

2. Estado de la Cuestión

2.1 Evaluación de la Usabilidad

El éxito de un programa depende tanto de la calidad del diseño de su interfaz, como del diseño de la interacción del usuario con el sistema. El usuario final solo juzga lo que ve y cómo se comporta el sistema frente a sus necesidades, de ahí la importancia de invertir dinero y tiempo para la evaluación y mejora de sus interfaces y de la interacción del usuario con el sistema [Hassan & Martín, 2004].

El objetivo de la evaluación de usabilidad es el de asegurar que las aplicaciones software evaluadas posean aspectos tales como eficacia, eficiencia y satisfacción de los usuarios [Alshamari & Mayhew, 2009].

En [Ivory & Hearst, 2001] se trata la evaluación de la usabilidad como un proceso en sí mismo, que requiere de un conjunto de actividades dependiendo del método empleado para su realización. Por ejemplo, implica recolectar datos de tiempo, errores, violaciones de las guías, entre otros, para interpretarlos y sugerir soluciones o mejoras para mitigar los problemas.

Uno de los factores más importantes que afectan a la evaluación de la usabilidad es la selección del método de evaluación de usabilidad, entre los cuales, [Nielsen, 1993] nos presenta una gran recopilación como la Evaluación Heurística, *Remote Testing*, *Thinking Aloud* y *Feature Inspection*. La selección de un método u otro depende de múltiples factores, dado que algunos de estos métodos requieren de recursos, como un completo laboratorio de usabilidad con espacios independientes para el desarrollo de las pruebas y tecnología específica, como cámaras de vídeo y equipos de observación [Perurena & Moráguez, 2013]. Otros factores son las tareas, la cantidad de usuarios, las características del evaluador y las medidas de usabilidad.

Muchas de las evaluaciones de usabilidad se llevan a cabo mediante pruebas de concepto o estudios de casos pero no siguen uno de los principales métodos empíricos que es el experimento controlado [Falessi et al., 2017; Ko et al., 2013, Wohlin et al., 2012]. Un experimento controlado es el medio por el cual se organiza y lleva a cabo el método científico clásico para identificar las relaciones causa-efecto. Concretamente en la Ingeniería del Software (IS), el experimento controlado es aquel en el cual las personas o grupos (unidades experimentales) llevan a cabo una o más tareas en aras de comparar diferentes poblaciones, procesos, métodos, técnicas, lenguajes o herramientas (tratamientos) [Sjøberg et al., 2005]. En nuestro experimento comparamos el desempeño de los sujetos (eficiencia, eficacia y satisfacción) ante la presencia y ausencia de determinados mecanismos de usabilidad en un sistema software. En este trabajo, otro método que utilizaremos es el meta-análisis, un procedimiento estadístico que combina los resultados a través de estudios que integran los resultados. Este tipo de evaluación es muy poderosa para determinar la usabilidad de un producto debido a que combina múltiples estudios para proveer apoyo cuantitativo más preciso [Sánchez, 2011].

2.2 Características de Usabilidad Funcionales

Desde la IS, en un principio, se consideraba que la usabilidad era un atributo que estaba únicamente relacionado con la interfaz de usuario (IU) y, por tanto, solamente repercutía en los requisitos no funcionales. Sin embargo, esta opinión ha cambiado debido a las conclusiones de distintos estudios. En [Juristo et al., 2007a] se obtuvieron, de forma empírica, evidencias que confirmaban la relación entre la usabilidad y el diseño del software. Aquí, se determinó también que el costo de repetir el trabajo para obtener un nivel aceptable de usabilidad sería mucho más alto del esperado de acuerdo a la hipótesis de separación entre la IU y el *core* del sistema software ya que, como estamos comentado, pueden existir atributos de usabilidad que afecten directamente a la funcionalidad del sistema.

En este trabajo se tendrá en cuenta la clasificación realizada en [Juristo et al., 2007b] donde se han identificado distintos tipos de Características de Usabilidad Funcionales agrupándolos en familias [Juristo et al., 2007b]. Cada tipo lleva el nombre de Mecanismo de Usabilidad (MU). En concreto, el MU Abortar Operación pertenece a la familia del Deshacer-Cancelar, el MU Retroalimentación de Progreso a la familia de Retroalimentación y el MU Preferencias a la familia de Perfil de Usuario. Así, los MUs representan funcionalidades particulares que pueden ser construidas dentro un sistema software para incrementar la usabilidad. Además, deben ser tratadas como requisitos funcionales y necesitan ser explícitamente especificados, al igual que cualquier otro requisito funcional. Si estas funcionalidades de usabilidad son descritas en el documento de especificación de requisitos, es más probable que sean incorporadas en el sistema software, contribuyendo a mejorar los niveles de usabilidad.

Para la realización de este trabajo, no se pretende tener en cuenta todas las características de usabilidad presentadas. Si así fuese, surgirían problemas importantes a la hora de identificar las variables experimentales y de construir los escenarios del experimento que se llevará a cabo.

Es por ello que existe la necesidad de determinar cuáles son las características más importantes para su estudio. En los trabajos de [Juristo et al., 2007a; Juristo et al., 2007b] se consideran que los tres mecanismos de usabilidad de mayor impacto en el diseño son: Abortar Operación, Retroalimentación de Progreso y Preferencias, por tanto, son los tres mecanismos considerados en esta investigación.

Los criterios que se utilizaron en los trabajos citados para llegar a esta conclusión han sido: nivel de impacto sobre el diseño en cuanto a número de funcionalidades que serían afectadas en las aplicaciones a desarrollar, facilidad de reconocimiento por parte de un usuario cuando use el sistema y facilidad de evaluación desde el punto de vista de las recomendaciones de la disciplina de la Interacción Persona-Ordenador.

2.3 Trabajos Relacionados

Como se afirma en [Fang & Holsapple, 2007], la psicología tiene una gran importancia en el área de la usabilidad. Se han propuesto dos tipos de técnicas de la usabilidad basadas en la psicología (técnica cognitiva y metáforas del pensamiento del ser humano) los cuales a diferencia de otros métodos como las heurísticas, tienen en cuenta los pensamientos de los sujetos en cada momento consiguiendo así, una mejora en la inspección de la usabilidad.

Otro tema relacionado es la evaluación de la usabilidad desde aspectos sintácticos abordados en [Castro et al., 2011], respecto principalmente a la forma de un sitio web concretando que una web debe ser capaz de guiar al usuario en los pasos a realizar en la aplicación sin que éste pierda nunca la perspectiva de la tarea.

Se nos presenta además una perspectiva innovadora de la usabilidad en el trabajo de [Bolchini et al., 2009], la usabilidad como herramienta para mejorar los valores corporativos percibidos por los consumidores, que no debe confundirse con la percepción de la calidad de la marca que es, únicamente, un subconjunto del primer grupo.

Además, considerando estudios experimentales, algunos trabajos están empezando a estudiar la usabilidad desde una perspectiva empírica. Por ejemplo, en [Panach et al., 2013] han investigado la incorporación de propiedades de usabilidad en un sistema software siguiendo un método de desarrollo dirigido por modelos (MDD). Panach y colaboradores realizan una validación experimental para comprobar si la usabilidad de las aplicaciones generadas con el nuevo método mejora o no tras el uso del software por determinados usuarios. El diseño experimental considera los atributos de eficiencia y satisfacción sobre una aplicación web de juguete que gestiona el alquiler de vehículos.

En otra investigación, [Aveledo et al., 2012] intentan comprobar los beneficios de la usabilidad de manera empírica estudiando el impacto de ciertas características de usabilidad en sistemas software desarrollados para un determinado contexto. El diseño experimental considera también una aplicación web de juguete pero evalúan los atributos de eficiencia, eficacia y satisfacción.

Dentro de la rama de agregación de datos, [Metzger, 2002], en su experimento, utiliza cuatro agregaciones de datos para calcular el nivel de desigualdad e inequidad en la salud de las poblaciones. En dicho experimento se concluye que cuando los índices se basan en modelos de regresión se recomienda utilizar datos agregados con el objetivo de que dichos modelos expliquen de forma robusta las series de datos, cumplan las restricciones de los modelos y fortalezcan los resultados.

A diferencia de los experimentos mencionados anteriormente, nuestra investigación aborda el estudio empírico de tres mecanismos de usabilidad, con alto impacto en el diseño [Juristo et al., 2007a]. Estudiamos el impacto de tales mecanismos en el contexto de un sistema web de gestión de artículos en línea con el fin de analizar el impacto real de la usabilidad más allá de un entorno de juguete. Además realizaremos agregaciones de datos y su análisis, ausente en la literatura, partiendo del experimento original [Bouza, 2016], la primera réplica realizada [Ferreira, 2017] y esta segunda réplica. De esta manera, obtendremos unos resultados más fiables y consistentes. Por un lado, en el experimento original hubo resultados concluyentes para la variable respuesta satisfacción pues la presencia de los tres mecanismos hizo que la satisfacción aumentara. La eficiencia nunca empeoraba cuando la aplicación presentaba alguno de los mecanismos pero tal y como se expone, en algunos casos se encontró que el factor orden de exposición de realización de las tareas era determinante. Por otro lado, en la primera réplica se concluye que los tres mecanismos de usabilidad tienen efecto significativo en la satisfacción del usuario tras haber interactuado con el sistema. También se aprecia cómo la ausencia de ciertos mecanismos de usabilidad como el caso del Abortar Operación y el Preferencias resulta casi determinante para la eficacia del usuario y se obtienen evidencias de que los tres

mecanismos de usabilidad afectan a la eficiencia del usuario en términos de rapidez. Por último, se menciona que en el caso de Abortar Operación se tienen efectos significativos en todos los atributos de calidad.

3. Experimento Verdadero

3.1 Objetivo, Investigación e Hipótesis

Nuestra investigación tiene como objetivo estudiar empíricamente los mecanismos de usabilidad: Abortar Operación (MU-ABR), Retroalimentación de Progreso (MU-PFB) y Preferencias (MU-PRF) en una aplicación web con el propósito de evaluar el impacto de la usabilidad con respecto a la eficiencia, eficacia y satisfacción desde la perspectiva de los usuarios en el contexto de usuarios no informáticos.

Conforme al objetivo, nuestro experimento trata de resolver la siguiente pregunta de investigación:

RQ: ¿La presencia de los mecanismos de usabilidad impacta en la usabilidad de la aplicación?

Más concretamente deseamos responder a las siguientes tres preguntas de investigación:

- RQ1: ¿La presencia del mecanismo de usabilidad Abortar Operación impacta en la usabilidad de la aplicación? Deseamos comprobar si la presencia de este mecanismo de usabilidad mejora o no la usabilidad de la aplicación en términos de la eficiencia, eficacia y satisfacción del usuario. La hipótesis nula que se encarga de esta pregunta de investigación es *H.1.X.0: No existe diferencia significativa en la EFICIENCIA | EFICACIA | SATISFACCIÓN del usuario al incorporar MU-ABR o al no incorporarlo*. Concretamente, esta hipótesis nula se desglosa en hipótesis nulas específicas, una por cada atributo de calidad (donde X representa 1. eficacia, 2. eficiencia y 3. satisfacción). Ellas son:
 - H.1.1.0: No existe diferencia significativa en la EFICIENCIA del usuario al incorporar el MU-ABR o al no incorporarlo.
 - H.1.2.0: No existe diferencia significativa en la EFICACIA del usuario al incorporar el MU-ABR o al no incorporarlo.
 - H.1.3.0: No existe diferencia significativa en la SATISFACCIÓN del usuario al incorporar el MU-ABR o al no incorporarlo.
- RQ2: ¿La presencia del mecanismo de usabilidad Retroalimentación de Progreso impacta en la usabilidad de la aplicación? Deseamos comprobar si la presencia de este mecanismo de usabilidad mejora o no la usabilidad de la aplicación en términos de la eficiencia, eficacia y satisfacción del usuario. La hipótesis nula que se encarga de esta pregunta de investigación es *H.2.X.0: No existe diferencia significativa en la EFICIENCIA | EFICACIA | SATISFACCIÓN del usuario al incorporar MU-PFB o al no incorporarlo*. Las hipótesis nulas específicas son:
 - H.2.1.0: No existe diferencia significativa en la EFICIENCIA del usuario al incorporar el MU-PFB o al no incorporarlo.
 - H.2.2.0: No existe diferencia significativa en la EFICACIA del usuario al incorporar el MU-PFB o al no incorporarlo.
 - H.2.3.0: No existe diferencia significativa en la SATISFACCIÓN del usuario al incorporar el MU-PFB o al no incorporarlo.

- RQ3: ¿La presencia del mecanismo de usabilidad MU-PRF impacta en la usabilidad de la aplicación? Deseamos comprobar si la presencia del mecanismo Preferencias mejora o no la usabilidad de la aplicación en términos de la eficiencia, eficacia y satisfacción del usuario. La hipótesis nula que se encarga de esta pregunta de investigación es *H.3.X.0: No existe diferencia significativa en la EFICIENCIA | EFICACIA | SATISFACCIÓN del usuario al incorporar el MU-PRF o al no incorporarlo*. Las hipótesis nulas específicas son:
 - H.3.1.0: No existe diferencia significativa en la EFICIENCIA del usuario al incorporar el MU-PRF o al no incorporarlo.
 - H.3.2.0: No existe diferencia significativa en la EFICACIA del usuario al incorporar el MU-PRF o al no incorporarlo.
 - H.3.3.0: No existe diferencia significativa en la SATISFACCIÓN del usuario al incorporar el MU-PRF o al no incorporarlo.

Concretamente, si las evidencias permiten rechazar la hipótesis nula, se entenderá aceptada la hipótesis alternativa, lo cual intenta revelar que cuando un mecanismo de usabilidad está presente o activado, se obtienen mejoras significativas en el atributo de calidad evaluado (eficiencia, eficacia o satisfacción).

3.2 Factores y Variables Respuesta

El factor objeto de nuestro experimento es el mecanismo de usabilidad. Dicho factor es la variable independiente que posee dos niveles: presencia y ausencia. Es decir, deseamos comparar el efecto de tener presente un mecanismo de usabilidad y no tenerlo mediante una tarea determinada. En el diseño experimental se estudiará paralelamente el impacto de tres mecanismos de usabilidad (MU-ABR, MU-PFB y MU-PRF), evaluando independientemente el efecto que produce cada uno.

En cuanto a las variables respuesta, se utilizarán tres medidas clásicas a la hora de analizar la usabilidad: la eficiencia, la eficacia y la satisfacción.

Para la variable eficiencia se medirán dos parámetros: el tiempo en segundos que utiliza cada sujeto para realizar cada una de las tareas y el número de clics que un sujeto necesita para completar la tarea, o dicho de otra manera, el nivel de interacción con el sistema para cada uno de los usuarios.

La variable eficacia se medirá con el porcentaje de tarea resuelta por un sujeto.

La variable satisfacción se medirá con el valor promedio de las preguntas de los cuestionarios post-tareas (Anexo B). Las preguntas tienen valores ordinales de la escala de Likert (desde uno = totalmente en desacuerdo al cinco = totalmente de acuerdo) [Sauro & Kindlund, 2005].

3.3 Contexto y Sujetos Experimentales

El experimento se lleva a cabo en un contexto académico con estudiantes de primer año de las carreras de Ciencias Contables, Ciencias Jurídicas, Ciencias del Deporte y Ciencias de la Salud de la Universidad Autónoma de Asunción de Paraguay. Todos ellos han participado de forma voluntaria como parte de la asignatura “Introducción a las TICs”, constituyendo un plus la participación en el mismo.

Los participantes están representados por personas ajenas a la informática, es decir, personas pertenecientes a otras áreas del conocimiento distintas a las ciencias de computación. Esta restricción forma parte del experimento, apoyando la idea de que sujetos que posean conocimientos informáticos limitados o nulos, pueden hacer uso de un sistema y reconocer los beneficios de la usabilidad. Además, de esta forma evitamos la influencia de usuarios informáticos que pueden estar familiarizados con este tipo de aplicaciones.

Para evitar el efecto de aprendizaje, no se ha proporcionado ningún tipo de sesión de información previa, simplemente, se les ha dado una visión general de la aplicación para que conozcan la estructura del experimento y puedan realizar las tareas con éxito.

La primera tarea que realizan es el cuestionario de familiaridad (Anexo A), del cual se destacan las siguientes características:

- La muestra final cuenta con 99 sujetos de los cuales 45 son hombres y 54 mujeres.
- La mayor concentración de participantes están entre 18-30, excepto nueve que pertenecerían a la franja entre 31-40 años y dos entre 41-50. Por las edades de los sujetos se infiere la presencia mayoritaria de estudiantes universitarios puesto que la ejecución experimental se llevó a cabo con estudiantes en las distintas carreras de la Universidad Autónoma de Asunción.
- Los sujetos experimentales son asiduos usuarios de Internet, pues han manifestado que lo usan en casa (73%), en el trabajo (19%) y un grupo menor en otros ámbitos (8%).
- En este mismo cuestionario se pregunta por las costumbres de los sujetos referidas a las compras en Internet. Mayoritariamente los participantes nunca han realizado compras en línea (66%). Un 15% lo ha hecho rara vez. Un 15% lo ha hecho a veces y un 4% con mayor frecuencia (siempre o casi siempre).

3.4 Diseño Experimental

El diseño del experimento siguiendo el proceso experimental propuesto en [Juristo & Moreno, 2010] consiste en un diseño *Between Subjects*. En este tipo de experimentos, cada sujeto experimental está asignado únicamente a un grupo que corresponde a una combinación de los niveles de los factores. Cada sujeto interactúa con los tres mecanismos de usabilidad, ya sea presente o ausente, mediante tareas asignadas aleatoriamente y ejecutadas de forma secuencial. De este modo, cada sujeto experimental proporciona un único valor para cada variable respuesta que será utilizado en el análisis estadístico.

Este diseño está caracterizado por tres matrices: tratamientos, orden de exposición y asignación de grupos. La matriz de tratamientos determina qué mecanismos de usabilidad estarán presentes y cuáles ausentes durante el experimento. En la Tabla 1 se presenta dicha matriz donde los ceros indican que el mecanismo de usabilidad está ausente y los unos indican que está presente. A modo de ejemplo, si un sujeto tiene asignado el tratamiento A realizará 3 tareas donde tendrá presente el MU-PRF en una de ellas.

Tratamiento	MU-ABR	MU-PFB	MU-PRF
A	0	0	1
B	1	0	0
C	0	1	0
D	1	1	1

Tabla 1: Matriz de Tratamiento

La segunda matriz es la de órdenes de exposición de los mecanismos, en esta matriz (Tabla 2) se establecen las secuencias posibles de realización de las tres tareas asociadas a cada uno de los mecanismos de usabilidad. Por ejemplo, si a un sujeto se le asigna el orden O1, éste seguirá una secuencia de acciones iniciando con la tarea del MU-ABR, luego con la tarea MU-PFB y finalmente con la tarea del MU-PRF.

Orden	Primero	Segundo	Tercero
O1	MU-ABR	MU-PFB	MU-PRF
O2	MU-ABR	MU-PRF	MU-PFB
O3	MU-PFB	MU-PRF	MU-ABR
O4	MU-PFB	MU-ABR	MU-PRF
O5	MU-PRF	MU-ABR	MU-PFB
O6	MU-PRF	MU-PFB	MU-ABR

Tabla 2: Matriz de Orden de Exposición

Finalmente, el diseño se completa con la última matriz que consiste en el producto cartesiano de la matriz de tratamientos con la matriz de órdenes. La resultante es la matriz de asignación de grupos y se describe en la Tabla 3. En total se tienen 24 grupos elegibles para un sujeto experimental. Como ejemplo concreto, si a un sujeto se le asigna dentro del GRUPO1 durante la ejecución experimental, dicho sujeto pertenece al tratamiento A ejecutando el orden O1. Es decir, el sujeto ejecutará primeramente la tarea con el MU-ABR ausente, a continuación la tarea con el MU-PFB ausente y finalmente la tarea con el MU-PRF presente.

Tratamiento	Orden	Grupo	Primero			Segundo			Tercero		
			MU-ABR	MU-PFB	MU-PRF	MU-ABR	MU-PFB	MU-PRF	MU-ABR	MU-PFB	MU-PRF
A	O1	GRUPO1	0				0				1
	O2	GRUPO2	0					1		0	
	O3	GRUPO3		0				1	0		
	O4	GRUPO4		0		0					1
	O5	GRUPO5			1	0				0	
	O6	GRUPO6			1		0		0		
B	O1	GRUPO7	1				0				0
	O2	GRUPO8	1					0		0	
	O3	GRUPO9		0				0	1		
	O4	GRUPO10		0		1					0
	O5	GRUPO11			0	1				0	
	O6	GRUPO12			0		0		1		
C	O1	GRUPO13	0				1				0
	O2	GRUPO14	0					0		1	
	O3	GRUPO15		1				0	0		
	O4	GRUPO16		1		0					0
	O5	GRUPO17			0	0				1	
	O6	GRUPO18			0		1		0		
D	O1	GRUPO19	1				1				1
	O2	GRUPO20	1					1		1	
	O3	GRUPO21		1				1	1		
	O4	GRUPO22		1		1					1
	O5	GRUPO23			1	1				1	
	O6	GRUPO24			1		1		1		

Tabla 3: Matriz de Asignación de Grupos

Se debe tener en cuenta que la matriz de tratamientos se usa para el proceso de análisis de datos mientras que la de órdenes de exposición tiene como fin atenuar cualquier efecto de aprendizaje. La matriz de asignación de grupos, se usa para asignar aleatoriamente un grupo a cada sujeto durante la ejecución del experimento.

3.5 Herramientas y Tareas

El instrumento de recolección de datos es una aplicación web. Dicha aplicación debe contar con características altamente interactivas, es decir, el sistema software debe depender de las acciones de un usuario para realizar una tarea. En este sentido, se utiliza una tienda en línea denominada *QuickStore* donde se incorporan los tres mecanismos de usabilidad MU-ABR, MU-PFB y MU-PRF. También, se dispone de la interfaz de administración del experimento [Ferreira & Acuña, 2017], la cual orientará al sujeto a lo largo de todo el experimento proporcionándole indicaciones de las actividades y acciones que debe realizar.

Adicionalmente, la interfaz de administración facilita el formateo y exportación de los datos empíricos recolectados (tiempos, clics, cuestionarios, etc.) para el análisis de impacto en los atributos de eficiencia, eficacia y satisfacción.

En el experimento se presentan las tres tareas que cada sujeto debe realizar. La tarea correspondiente al mecanismo de Preferencias está desglosada en dos sub-tareas. Ellas son:

- **Tarea Abortar Operación:** el sujeto simulará una operación de cancelar durante el procesamiento de su carrito de compra. Al ingresar al sistema, el carrito del usuario ya contendrá varios artículos. La tarea consiste en que el sujeto ingrese a su carrito de compra y modifique sus datos (por ejemplo, incrementar la cantidad de cualquiera de los artículos, completar un código promocional, etc.) y, seguidamente, cancele la operación. Si el mecanismo de usabilidad está presente, el usuario dispondrá de una opción de cancelación rápida, el cual, únicamente, solicitará confirmación sobre los cambios pendientes. Por el contrario, si el mecanismo de usabilidad no está activo la única forma de realizar la cancelación es la manual, es decir, debe deshacer los cambios que realizó desde el inicio de la tarea. En la Figura 9 del Anexo B se muestra el diseño de la tarea propuesta para el MU-ABR.
- **Tarea Retroalimentación de Progreso:** en esta segunda tarea, el sujeto simulará una operación de búsqueda de un artículo, en concreto de un libro, y una vez localizado, incorporarlo al carrito de compra. El sujeto inicia la tarea desde la pantalla principal de la aplicación *QuickStore*, en la que puede realizar una búsqueda con los criterios que estime, por ejemplo, el nombre del artículo. Si la búsqueda es exitosa, únicamente debe pulsar en el botón Añadir al Carrito. Si el mecanismo de usabilidad está activo, mientras se realiza la búsqueda se mostrará una barra de progreso informando al usuario de que la acción se está ejecutando con un mensaje, al final, indicando la cantidad de artículos encontrados. Si el mecanismo de usabilidad no está presente, se ejecutará igualmente la búsqueda, pero no se informará al usuario de que la acción se está realizando. En la Figura 10 del Anexo B se muestra el diseño de la tarea propuesta para el MU-PFB.
- **Tarea Preferencias:** esta tarea está dividida en dos partes. En primer lugar se realizará la tarea preferencias básica y, seguidamente, la tarea preferencias ficticia. En la Figura 11 del Anexo B se muestra el diseño de ambas tareas para el MU-PRF.
 - **Tarea Preferencias Normal o Básica:** el sujeto intentará personalizar la interfaz de la aplicación. Inicialmente, el aspecto de la interfaz no será agradable, pues el tamaño de la letra será pequeño y la tipografía será poco legible. Si el mecanismo de usabilidad está activo, el usuario podrá personalizar ciertos aspectos de la tienda a su gusto. Por el contrario, si el mecanismo no está presente el usuario no podrá modificar la apariencia de la interfaz desde la propia aplicación.
 - **Tarea Preferencias Ficticia:** en esta tarea se solicita al usuario buscar cierta información, proporcionada en la aplicación, relacionada con los plazos de devolución de los artículos comprados. Si el sujeto ha realizado la modificación de la interfaz del sistema, podrá ver fácilmente dónde está el enlace para encontrar la información que se propone. Sin embargo, si el usuario no ha sido capaz, o no ha querido, modificar la apariencia de la aplicación, tendrá una interfaz con alta dificultad para encontrar la información requerida.

El conjunto de estas tareas principales conforman el escenario de ejecución del experimento. La realización de cada tarea supone que el sujeto es un cliente de la tienda, es decir, interactúa con la interfaz final del sistema haciendo tareas no administrativas. El orden de ejecución es aleatorio y está determinado por la matriz de asignación de grupos (Tabla 3).

Finalmente, los sujetos deben completar el cuestionario de familiaridad de forma previa a la realización de las tareas propuestas (Anexo A). Además, por cada tarea realizada el sujeto debe completar un cuestionario de satisfacción post-tarea correspondiente al mecanismo utilizado. Los cuestionarios de satisfacción post-tareas se encuentran en el Anexo B.

3.6 Operación

Los sujetos experimentales no tenían conocimiento del objetivo del estudio ni de las hipótesis de la investigación. No se ha mencionado la palabra “Experimento” para evitar cualquier efecto negativo y por el contrario, hemos referido al experimento como “Evaluación”. También hemos informado que los resultados de dicha evaluación servirán para mejorar ciertos aspectos en las aplicaciones web y se garantiza la confidencialidad de los mismos.

Finalmente, hemos notificado que la participación es voluntaria y en caso de que el sujeto acepte se le anima a invertir su esfuerzo en la realización de las tareas. No hemos proporcionado ningún material adicional más que el enlace al que deben acceder para iniciar la evaluación.

El experimento se ejecutó durante 5 meses, entre noviembre y marzo del año 2017-2018 en la Universidad Autónoma de Asunción utilizando la plataforma de educación a distancia. La ejecución llevada a cabo no ha interferido con los objetivos del curso de cada carrera ya que la misma ha sido propuesta como una tarea de desafío opcional a la que los estudiantes acceden libre y voluntariamente. Todos los sujetos son asignados aleatoriamente a un grupo del diseño experimental. De este modo se espera evitar cualquier influencia resultado del efecto de aprendizaje. En este experimento, un grupo determina el orden de realización de las tareas y qué mecanismos de usabilidad estarán presentes o ausentes.

3.7 Amenazas a la Validez

En este apartado se introduce la validez del experimento y sus distintas amenazas, tanto de carácter interno como externo.

Un experimento resulta válido en la medida que los resultados pueden ser atribuidos inequívocamente a la variable independiente [van Welie, 2008]. La validez interna se articula en torno a la supervisión del proceso en aras de establecer las relaciones de asociación entre las variables independientes y la dependiente. La validez externa está relacionada con el establecimiento de las condiciones que permiten la generalización de los resultados al ámbito natural en el que aparecen los procesos investigados [Acuña et al., 2008].

3.7.1 Validez Interna

La validez interna se refiere a cuánta confianza tenemos en los resultados, en su validez y en poder interpretarlos.

Esta validez se relaciona con la calidad del experimento y se logra cuando hay control, sin embargo, las amenazas a la validez interna son influencias que pueden afectar a la variable independiente con respecto a la causalidad fuera del conocimiento del investigador. En nuestro experimento en concreto, encontramos las siguientes amenazas a la validez interna:

1. Conocimiento de la tecnología: aunque todos los sujetos parten de un nivel novato de experiencia hacia este tipo de experimentos, no todos los sujetos tienen el mismo nivel de conocimiento de la actividad que se va a desarrollar. Además, en el cuestionario de familiaridad se desprende un alto porcentaje de conocimiento de páginas web pero un bajo porcentaje de compra en línea.
2. Orden de realización de las tareas: en el caso en el que las tareas de un mismo mecanismo se ejecuten secuencialmente, podría producirse un sesgo por efecto de aprendizaje.
3. Baja experiencia de los usuarios: tal y como hemos comentado anteriormente, todos los participantes parten de un nivel novato en este tipo de experimentos y en esta aplicación en concreto. Por ello, es posible que no entiendan bien cómo realizar las tareas, que no presten atención a las indicaciones, entre otras cosas.
4. Motivación: es de esperar que no todos los sujetos afronten las tareas de la misma manera, especialmente si están realizando otras cosas al mismo tiempo.

Para mitigar cualquier posible influencia de las dos primeras amenazas los sujetos son asignados aleatoriamente dentro de un grupo manteniendo el balance de participantes entre grupos. Este proceso de aleatorización según [Cochran & Cox, 1980] es una precaución contra interferencias y es aconsejable tomarse el tiempo de aleatorizar, aun cuando no se espere que haya un sesgo importante al no hacerlo.

La tercera amenaza se suple introduciendo el orden como un factor dentro del diseño. Además, en una sesión, los profesores les dan unas indicaciones previas para que se familiaricen con la interfaz.

Finalmente, la cuarta amenaza no la podemos evitar, sin embargo, a los efectos de atenuar su impacto, la realización de la “evaluación” constituía un plus en la nota de la asignatura.

3.7.2 Validez Externa

La validez externa hace referencia a la extensión y forma en que los resultados pueden generalizarse en otros contextos.

Consideramos que los resultados del experimento no son generalizables a todos los usuarios ya que los participantes seleccionados son no informáticos para evitar el sesgo por la familiaridad con la tecnología. Adicionalmente, estos sujetos son personas que conforman una parte importante de la población de usuarios que hacen poco uso de aplicaciones web para la compra de productos y que permiten obtener evidencias empíricas con buena confianza sobre el impacto de los mecanismos de usabilidad analizados a nivel de usuarios no informáticos.

4. Enfoque de Análisis

4.1 Análisis de Datos

Dividimos el análisis en tres diferentes apartados (uno por cada mecanismo de usabilidad: Abortar Operación, Retroalimentación de Progreso y Preferencias). En cada apartado evaluaremos el efecto del mecanismo de usabilidad en todas las variables respuesta (número de clics, tiempo, porcentaje de cumplimiento de tareas y satisfacción).

Dentro de cada apartado seguimos el mismo proceso: Mostraremos las distribuciones de las puntuaciones de cada variable respuesta por medio de diagramas de violín. Se comparan dos grupos: el grupo donde estaba presente el mecanismo de usabilidad y el grupo donde el mecanismo no estaba presente. En el Anexo C se pueden encontrar los respectivos diagramas de caja. Los estadísticos descriptivos (mínimo, máximo, media, mediana, rango intercuartílico, cuartiles del 1 al 4, desviación típica, varianza, *skewness* y *kurtosis*) se proporcionan para complementar los diagramas de violín y de caja. A continuación, se realiza la significación estadística donde se calcula el p-valor de la prueba U-Mann Whitney, el tamaño de efecto Cliff's delta, la evaluación de la significación estadística, evaluación del tamaño de efecto y probabilidad de superioridad de un grupo sobre otro para cada variable respuesta y mecanismo. Para la realización de todo el estudio estadístico se ha utilizado el lenguaje R [Field et al., 2012].

Primero, el análisis del mecanismo Abortar Operación se encuentra en el apartado 4.1.1. A continuación el análisis del mecanismo Retroalimentación de Progreso en el apartado 4.1.2 y por último, el análisis del mecanismo Preferencias en el apartado 4.1.3.

4.1.1 RQ1: Abortar Operación (MU-ABR)

Con respecto a la **Eficiencia**, vista como el **grado de interacción**, podemos observar (Tabla 4) como la media y la mediana son menores en los casos en los que el mecanismo de usabilidad está presente. Lo mismo ocurre con la desviación típica, tiende a ser menor para los sujetos quienes tenían el mecanismo presente frente a aquellos quienes no lo tenían. Respecto al valor mínimo nulo nos indica que los sujetos no han interactuado con el sistema, por lo que descartamos dichos valores. Por último, observamos que los valores de *skewness* y *kurtosis* son ambos positivos.

Analizando la **Eficiencia**, considerando la **variable tiempo**, notamos en la Tabla 4, como nuevamente la media, la mediana y la desviación típica son menores en los casos en los que el mecanismo está presente frente a cuando está ausente. Además, el índice de asimetría es positivo en ambos casos mientras que el índice de curtosis es positivo para la presencia pero negativo en el caso de la ausencia.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Clics Presencia	0	32	12,49	12	8,5	8	12	16,5	32	6,93	48,01	0,47	0,04
Clics Ausencia	0	69	17,13	15	12,5	9,5	15	22	69	12,12	146,94	1,68	4,99
Tiempo Presencia	14,89	433,05	122,37	103,64	64,16	76,18	103,64	140,15	433,05	77,17	5954,74	1,74	3,80
Tiempo Ausencia	2,38	501,45	191,61	174,02	197,18	81,15	174,02	278,33	501,45	123,54	15261,35	0,56	-0,35

Tabla 4: Estadísticos descriptivos para la Eficiencia en el Mecanismo de Usabilidad Abortar Operación

En cuanto a la medida de **Eficacia**, la media, la mediana y la desviación típica son mayores cuando el mecanismo está presente (Tabla 5). En cuanto al índice de curtosis, es positivo cuando el mecanismo está presente y cuando no lo está mientras que el índice de asimetría es negativo en ambos casos.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Porcentaje Presencia	0	100	86,08511	100	12,5	87,5	100	100	100	28,85902	832,8431	-2,139198	3,452301
Porcentaje Ausencia	0	75	68,08511	75	0	75	75	75	75	19,29612	372,3404	-2,834846	6,97324

Tabla 5: Estadísticos descriptivos para la Eficacia en el Mecanismo de Usabilidad Abortar Operación

Finalmente analizamos la variable **Satisfacción** observando la Tabla 6. Los valores de la media y la mediana son mayores cuando el mecanismo está presente frente a cuando está ausente. Sin embargo, la desviación típica es ligeramente menor en este caso. En cuanto al índice de asimetría es negativo cuando el mecanismo está presente y positivo cuando no mientras que con el índice de curtosis ocurre a la inversa.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Satisfacción Presencia	1	5	3,96	4	1,5	3,5	4	5	5	1,02	1,05	-1,20	1,12
Satisfacción Ausencia	1	5	2,90	3	2	2	3	4	5	1,30	1,70	0,04	-1,32

Tabla 6: Estadísticos descriptivos para la Satisfacción en el Mecanismo de Usabilidad Abortar Operación

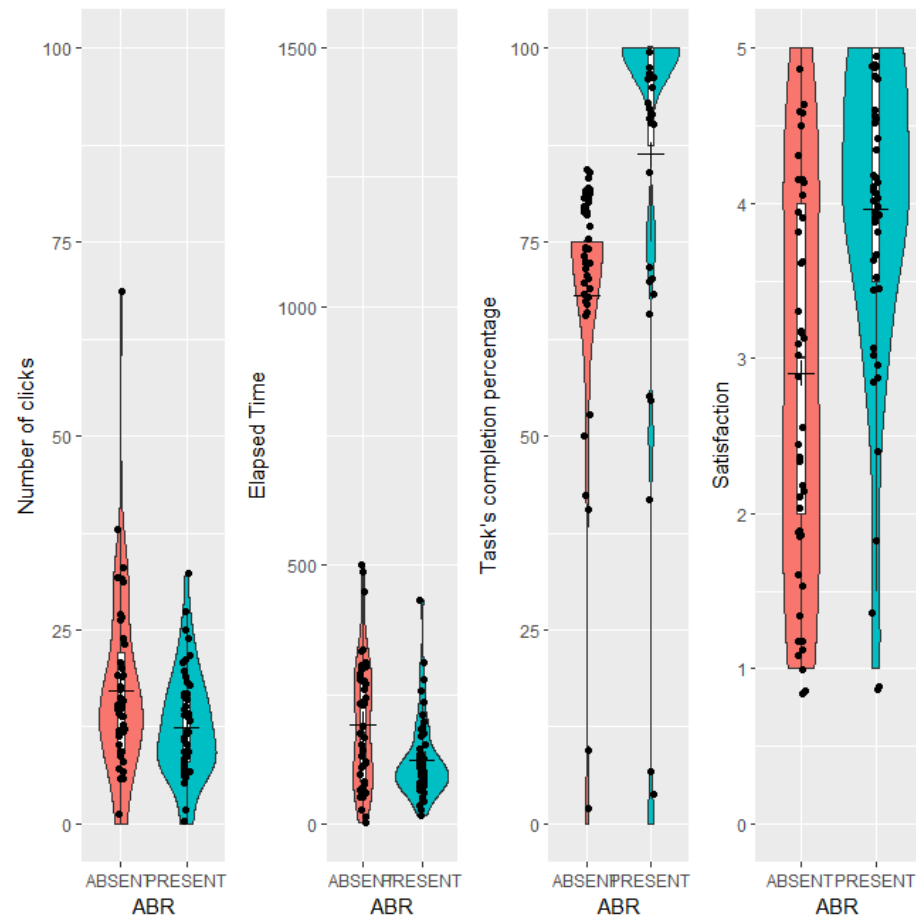


Figura 2: Diagrama de violín para el Mecanismo de Usabilidad Abortar Operación

Como podemos observar en la Figura 2, la gráfica correspondiente al número de clics y la del tiempo son bastante similares. En ambas están bastante más separados cuando el mecanismo de usabilidad está ausente que cuando está presente. Siendo además algo mayor en el caso de ausencia del mecanismo Abortar Operación.

El porcentaje de completitud en ambos casos varía por toda la gráfica pero estando la mayor concentración en el caso de ausencia en torno al 75% y en el caso de presencia en torno al 100%.

La satisfacción tiene su grueso en torno al 4-5 en el caso de presencia, mientras que en el caso de ausencia observamos que los valores están más separados.

La Tabla 7 muestra la significación estadística (p-valor de la prueba U-Mann Whitney), tamaño de efecto (Cliff's delta), evaluación de la significación estadística, evaluación del tamaño de efecto y probabilidad de superioridad de un grupo sobre otro para cada variable respuesta para el mecanismo de Abortar Operación.

Variable Respuesta	p-valor	Tamaño de efecto (Cliff's delta)	Evaluación p-valor	Evaluación de Tamaño de efecto	Pr (Superioridad)	
					(Presente>Ausente)	(Ausente>Presente)
Clic	0,042	0,24	Significativo	Pequeño	0,36	0,598
Tiempo	0,007	0,32	Significativo	Pequeño	0,34	0,659
Porcentaje	<0,001	-0,64	Significativo	Grande	0,766	0,123
Satisfacción	<0,001	-0,46	Significativo	Medio	0,675	0,212

Tabla 7: Datos estadísticos del Mecanismo de Usabilidad Abortar Operación

Los resultados que se observan son los siguientes (Tabla 7):

- Existe una probabilidad del 36% de que los sujetos con acceso al mecanismo de Abortar Operación hagan clic más veces que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 59,8% de probabilidad de que los sujetos con acceso al mecanismo de Abortar Operación hagan clic menos veces que aquellos que no tienen acceso a dicho mecanismo. La diferencia es estadísticamente significativa y el tamaño de efecto es pequeño.
- Hay un 34% de probabilidad de que los sujetos con acceso al mecanismo de Abortar Operación pasen más tiempo que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 65,9% de probabilidad de que los sujetos con acceso al mecanismo de Abortar Operación pasen menos tiempo que aquellos que no tienen acceso a dicho mecanismo. La diferencia es estadísticamente significativa y el tamaño de efecto es pequeño.
- Existe un 76,6% de probabilidad de que los sujetos con acceso al mecanismo de Abortar Operación sean más eficaces que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 12,3% de probabilidad de que los sujetos con acceso al mecanismo de Abortar Operación sean menos eficientes que aquellos que no tienen acceso a dicho mecanismo. La diferencia es estadísticamente significativa y el tamaño de efecto es grande.
- Existe un 67,5% de probabilidad de que los sujetos con acceso al mecanismo de Abortar Operación estén más satisfechos que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 21,2% de probabilidad de que los sujetos con acceso al mecanismo de Abortar Operación estén menos satisfechos que aquellos que no tienen acceso a dicho mecanismo de usabilidad. La diferencia es estadísticamente significativa y el tamaño de efecto es medio.

4.1.2 RQ2: Retroalimentación de Progreso (MU-PFB)

Con respecto a la **Eficiencia**, vista como el **grado de interacción** con el sistema, observamos en la Tabla 8 como la media, la mediana y la desviación típica con menores en el caso de la presencia del mecanismo de usabilidad. Además, notamos en este caso un valor mínimo nulo, lo cual representa una falta de interacción con el sistema y debe ser eliminado para la investigación. Por último destacamos que tanto el índice de asimetría como el de curtosis son ambos positivos.

Analizando la **Eficiencia** desde la **variable tiempo** (Tabla 8) observamos como la media, la mediana y la desviación típica son menores en el caso de la presencia frente a la ausencia. Además, los índices de simetría y de curtosis con positivos.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Kurtosis
Clicks Presencia	0	17	4,49	4	3	3	4	6	17	2,92	8,53	1,87	25,32
Clicks Ausencia	2	76	8,79	5	6	4	5	10	76	10,82	117,05	4,63	5,67
Tiempo Presencia	2,80	250,14	80,41	65,30	42,64	47,72	65,30	90,04	250,14	54,79	3002,02	1,38	1.31
Tiempo Ausencia	19,85	603,71	103,05	68,70	57,94	45,40	68,70	103,34	603,71	115,31	13296,41	2,99	9,30

Tabla 8: Estadísticos descriptivos para la Eficiencia en el Mecanismo de Usabilidad Retroalimentación de Progreso

A la hora de analizar la **Eficacia**, observamos los datos descriptivos del porcentaje de tarea realizada por cada sujeto en la Tabla 9. Éstos nos muestran datos muy alternos, la media es menor en el caso de la presencia, la mediana es exactamente la misma para los dos casos y la desviación típica es mayor en el caso de la presencia. Por otro lado, el índice de asimetría es negativo tanto cuando el mecanismo está presente como cuando no mientras que el índice de curtosis es positivo en ambos casos.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Porcentaje Presencia	0	100	93,33	100	0	100	100	100	100	25,23	636,36	-3,36	9,50
Porcentaje Ausencia	0	100	98,11	100	0	100	100	100	100	13,74	188,68	-6,87	46,11

Tabla 9: Estadísticos descriptivos para la Eficacia en el Mecanismo de Usabilidad Retroalimentación de Progreso

Por último analizando la **Satisfacción**, por un lado podemos ver (Tabla 10) como la media es mayor en el caso de la presencia, mientras que la mediana y la desviación típica son menores en este caso. Por otro lado, observamos como el mínimo y el máximo envuelven todo el rango de valores. Además, el índice de asimetría es negativo para ambos casos y el de curtosis es negativo para el caso de ausencia del mecanismo.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Satisfacción Presencia	1	5	4,04	4	1,5	3,5	4	5	5	0,97	0,94	-1,24	1,82
Satisfacción Ausencia	1	5	3,96	4,5	2	3	4,5	5	5	1,09	1,18	-0,71	-0,58

Tabla 10: Estadísticos descriptivos para la Satisfacción en el Mecanismo de Usabilidad Retroalimentación de Progreso

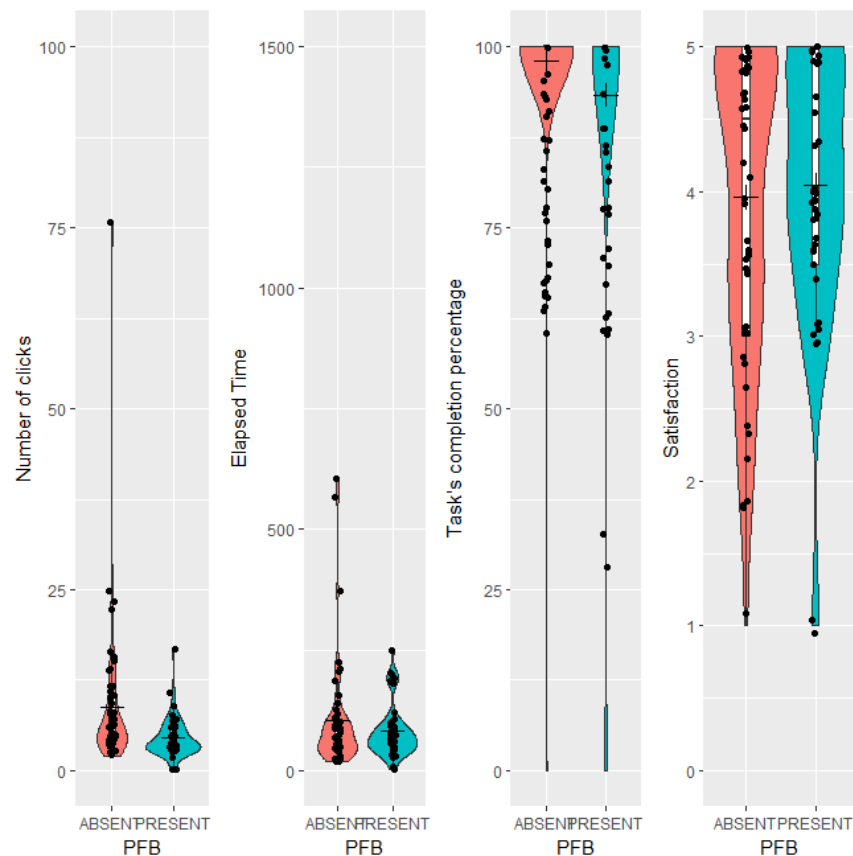


Figura 3: Diagrama de violín para el Mecanismo de Usabilidad Retroalimentación de Progreso

En la Figura 3, en la que en el número de clics y en el tiempo, notamos el grueso en ambos casos en torno a los números bajos pero en el caso de la ausencia, los valores se extienden un poco más a los valores más altos de la gráfica.

El porcentaje de completitud de la tarea es muy parecido, incluso algo mejor cuando el mecanismo de usabilidad está ausente. La satisfacción es también muy parecida, estando los valores de ausencia del mecanismo más separados mientras que los de presencia del mecanismo prácticamente todos rondan los valores 4-5.

La Tabla 11 muestra la significación estadística (p-valor de la prueba U-Mann Whitney), tamaño de efecto (Cliff's delta), evaluación de la significación estadística, evaluación del tamaño de efecto y probabilidad de superioridad de un grupo sobre otro para cada variable de respuesta para el mecanismo de Retroalimentación de Progreso.

Variable Respuesta	p-valor	Tamaño de efecto (Cliff's delta)	Evaluación p-valor	Evaluación tamaño de efecto	Pr (Superioridad)	
					(Presente>Ausente)	(Ausente>Presente)
Clic	<0,001	0,43	Significativo	Medio	0,234	0,66
Tiempo	0,679	0,05	No significativo	Despreciable	0,475	0,525
Porcentaje	0,24	0,05	No significativo	Despreciable	0,018	0,065
Satisfacción	0,809	-0,03	No significativo	Despreciable	0,417	0,389

Tabla 11: Datos estadísticos del Mecanismo de Usabilidad Retroalimentación de Progreso

Los resultados obtenidos son los siguientes (Tabla 11):

- Hay un 23,4% de probabilidad de que los sujetos con acceso al mecanismo de Retroalimentación de Progreso hagan clic más veces que aquellos que no tienen acceso a dicho mecanismo. En otras palabras, hay un 66% de probabilidad de que los sujetos con acceso al mecanismo de Retroalimentación de Progreso hagan clic menos veces que aquellos que no tienen acceso a dicho mecanismo. La diferencia es estadísticamente significativa y el tamaño de efecto es medio.
- Existe una probabilidad del 47,5% de que los sujetos con acceso al mecanismo de Retroalimentación de Progreso pasen más tiempo para realizar la tarea que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 52,5% de probabilidad de que los sujetos con acceso al mecanismo pasen menos tiempo que aquellos que no tienen acceso a dicho mecanismo. La diferencia no es estadísticamente significativa y el tamaño de efecto es despreciable.
- Existe una probabilidad del 1,8% de que los sujetos con acceso al mecanismo de Retroalimentación de Progreso sean más eficientes que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 6,5% de probabilidad de que los sujetos con acceso al mecanismo sean menos eficientes que aquellos que no tienen acceso a dicho mecanismo. La diferencia no es estadísticamente significativa y el tamaño de efecto es despreciable.
- Existe un 41,7% de probabilidad de que los sujetos con acceso al mecanismo de Retroalimentación de Progreso estén más satisfechos que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 38,9% de probabilidad de que los sujetos con acceso al mecanismo de Retroalimentación de Progreso estén menos satisfechos que aquellos que no tienen acceso a dicho mecanismo. La diferencia no es estadísticamente significativa y el tamaño de efecto es despreciable.

4.1.3 RQ3: Preferencias (MU-PRF)

Con respecto a la **Eficiencia**, vista como **el grado de interacción** del usuario con el sistema, en la Tabla 12 observamos como la media, la mediana y la desviación típica son menores en el caso en el que mecanismo está presente frente al caso en el que está ausente. Al igual que ocurría en el caso de Abortar Operación el valor mínimo es cero, lo cual representa a aquellos sujetos que no han llegado a interactuar con el sistema y que por tanto, no incluiremos en nuestra investigación. En cuanto al índice de asimetría y el de curtosis es positivo en ambos casos.

Analizando la **Eficiencia** desde la **variable tiempo** (Tabla 12), vemos como nuevamente la media, la mediana y la desviación típica son menores en el caso de presencia. Además, el índice de asimetría y de curtosis son positivos en ambas situaciones.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Clics Presencia	0	50	5,94	3,5	4	2	3,5	6	50	8,52	72,55	3,37	13,08
Clics Ausencia	0	53	10,48	5,5	12	1	5,5	13	53	13,05	170,43	1,48	1,39
Tiempo Presencia	11,19	337,42	88,86	60,01	86,51	33,90	60,01	120,40	337,42	76,65	5874,92	1,44	1,83
Tiempo Ausencia	2,18	590,15	119,97	71,65	133,48	30,75	71,65	164,23	590,15	137,81	18992,88	1,66	2,01

Tabla 12: Estadísticos descriptivos para la Eficiencia en el Mecanismo de Usabilidad Preferencias

En cuanto a la medida de **Eficacia** que medimos mediante el porcentaje de tarea realizada, observamos (Tabla 13) como la media y la mediana son claramente superiores mientras que la desviación típica es ligeramente menor en el caso de la presencia frente a la ausencia del mecanismo de usabilidad. Además, el índice de asimetría y de curtosis es negativo en el caso de presencia y en el caso de ausencia el índice de asimetría es positivo pero el de curtosis es negativo.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Porcentaje Presencia	0	100	72	100	100	0	100	100	100	45,36	2057,14	-0,95	-1,12
Porcentaje Ausencia	0	100	29,17	0	100	0	0	100	100	45,93	2109,93	0,89	-1,24

Tabla 13: Estadísticos descriptivos para la Eficacia en el Mecanismo de Usabilidad Preferencias

Finalmente, analizando la **Satisfacción** en la Tabla 14, la media y la mediana son superiores en el caso de la presencia del mecanismo a diferencia de la desviación típica. Los índices de asimetría y de curtosis son negativos en el caso en el que el mecanismo está presente y, en el caso contrario el índice de asimetría es positivo pero el de curtosis no.

Métrica	Mínimo	Máximo	Media	Mediana	IQR	Q1	Q2	Q3	Q4	Desviación Típica	Varianza	Skewness	Curtosis
Satisfacción Presencia	1	5	3,89	4	2	3	4	5	5	1,18	1,40	-0,85	-0,39
Satisfacción Ausencia	1	5	2,43	2	1,13	1,88	2	3	5	1,20	1,43	0,96	-0,078

Tabla 14: Estadísticos descriptivos para la Satisfacción en el Mecanismo de Usabilidad Preferencias

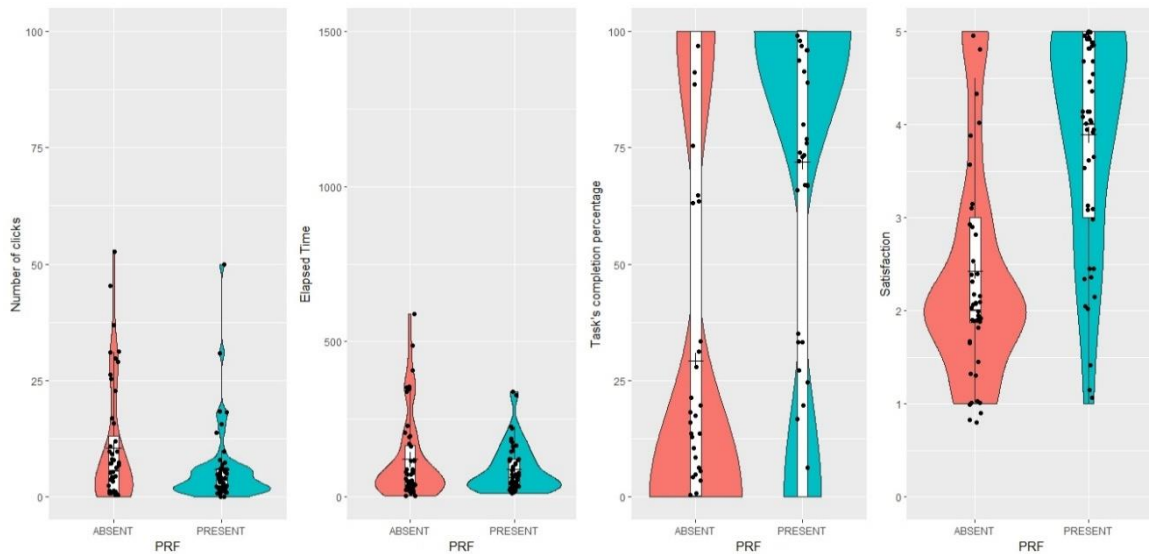


Figura 4: Diagrama de violín para el Mecanismo de Usabilidad Preferencias

Por último observando la Figura 4, se puede ver que tanto la gráfica del número de clics como la del tiempo son muy similares, estando más separados los valores de ausente y siendo en ambos casos mayores. En cuanto a la completitud, notamos gráficas muy parecidas pero a la inversa, estando el grueso de la gráfica en torno al 100% en el caso de presencia y en torno a 0% en el caso contrario.

La satisfacción es superior en el caso en el que el mecanismo está presente, con la mayoría de los valores en torno al 4-5, mientras que en el caso de ausencia, la mayoría ronda el valor 2.

La Tabla 15 muestra la significación estadística (p-valor de la prueba U-Mann Whitney), tamaño de efecto (Cliff's delta), evaluación de la significación estadística, evaluación del tamaño de efecto y probabilidad de superioridad de un grupo sobre otro para cada variable respuesta para el mecanismo de Preferencias.

Variable Respuesta	p-valor	Tamaño de efecto (Cliff's delta)	Evaluación p-valor	Evaluación tamaño de efecto	Pr (Superioridad)	
					(Presente>Ausente)	(Ausente>Presente)
Clic	0,304	0,12	No significativo	Despreciable	0,41	0,531
Tiempo	0,828	0,03	No significativo	Despreciable	0,487	0,413
Porcentaje	<0,001	-0,43	Significativo	Medio	0,51	0,082
Satisfacción	<0,001	-0,59	Significativo	Grande	0,75	0,155

Tabla 15: Datos estadísticos del Mecanismo de Usabilidad Preferencias

Los resultados obtenidos son los siguientes (Tabla 15):

- Hay un 41% de probabilidad de que los sujetos con acceso al mecanismo de Preferencias hagan clic más veces que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay una probabilidad del 53,1% de que los sujetos con acceso al mecanismo de Preferencias hagan clic menos veces

que aquellos que no tienen acceso a dicho mecanismo de usabilidad. La diferencia no es estadísticamente significativa y el tamaño de efecto es despreciable.

- Hay un 48,7% de probabilidad de que los sujetos con acceso al mecanismo de Preferencias pasen más tiempo que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 41,3% de probabilidad de que los sujetos con acceso al mecanismo de Preferencias pasen menos tiempo que aquellos que no tienen acceso a dicho mecanismo de usabilidad. La diferencia no es estadísticamente significativa y el tamaño de efecto es despreciable.
- Existe una probabilidad del 51% de que los sujetos con acceso al mecanismo de Preferencias sean más eficientes que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, hay un 8,2% de probabilidad de que los sujetos con acceso al mecanismo de Preferencias sean menos eficientes que aquellos que no tienen acceso a dicho mecanismo de usabilidad. La diferencia es estadísticamente significativa y el tamaño de efecto es medio.
- Existe un 75% de probabilidad de que los sujetos con acceso al mecanismo de Preferencias estén más satisfechos que aquellos que no tienen acceso a dicho mecanismo de usabilidad. En otras palabras, existe una probabilidad del 15,5% de que los sujetos con acceso al mecanismo de Preferencias estén menos satisfechos que aquellos que no tienen acceso a dicho mecanismo de usabilidad. La diferencia es estadísticamente significativa y el tamaño de efecto es grande.

4.2 Discusión de Resultados

En esta sección, discutimos los resultados cuantitativos para responder a las preguntas de investigación. En la Tabla 16 se resume los resultados del experimento. El símbolo * en la columna de Test denota la existencia de diferencias estadísticamente significativas, mientras que el símbolo – denota diferencias irrelevantes. El tamaño de efecto está descrito entre corchetes.

MU	Eficiencia				Eficacia		Satisfacción	
	Clics		Tiempo		Porcentaje		Valor	
	Hipótesis	Test	Hipótesis	Test	Hipótesis	Test	Hipótesis	Test
MU-ABR	H.1.1.0	*[Pequeño]	H.1.1.0	*[Pequeño]	H.1.2.1	*[Grande]	H.1.3.0	*[Medio]
MU-PFB	H.2.1.0	*[Medio]	H.2.1.0	-	H.2.2.0	-	H.2.3.0	-
MU-PRF	H.3.1.0	-	H.3.1.0	-	H.3.2.1	*[Medio]	H.3.3.1	*[Grande]

Tabla 16: Resumen de los resultados Experimentales

4.2.1 RQ1: Abortar Operación (MU-ABR)

Hemos encontrado que la presencia de este mecanismo de usabilidad tiene efecto en los tres atributos de calidad. Considerando la **Eficiencia**, las diferencias son significativas entendiéndose como que los sujetos con acceso al mecanismo mejoran el tiempo y los clics que los que no tienen acceso a dicho mecanismo. Creemos que al no existir una opción de cancelación directa los sujetos parecen invertir más tiempo y más cantidad de clics en realizar la tarea. Sin embargo, aunque existan diferencias significativas en los clics, el tamaño de efecto es pequeño. Esto que hace que dicha diferencia sea irrelevante y la mejora a nivel de interacción sea pobre (~ 5 clics). En cuanto al tiempo, aunque las

diferencias sean significativas, el tamaño de efecto es pequeño por lo que tampoco parece ser tan relevante y entendemos que la mejora en la rapidez del usuario es mínima. En ambos casos, podemos aceptar la hipótesis nula H.1.1.0: No existe diferencia significativa en la EFICIENCIA del usuario al incorporar el MU-ABR o al no incorporarlo. Sin embargo, nótese que hay una mayor probabilidad de que la presencia del MU-ABR tienda a mejorar la eficiencia en términos de rapidez y grado de interacción del usuario.

Las evidencias se acentúan con mayor intensidad hacia una mejora en la **Eficacia**. Los resultados confirman que existe diferencia significativa en la Eficacia del usuario. Comparando las medias, los sujetos han logrado cumplir en mayor porcentaje con el desafío de la tarea gracias al mecanismo presente. En cuanto a la **Satisfacción**, las diferencias son significativas entendiéndose como que los sujetos con acceso al mecanismo están más satisfechos que los que no tienen acceso a dicho mecanismo. Sin embargo, aunque existan diferencias significativas en la satisfacción, el tamaño de efecto es medio. Esto hace que dicha diferencia sea irrelevante y la mejora a nivel de satisfacción sea pobre (~1 punto). En el caso de la Eficacia, además de que las diferencias son estadísticamente significativas, el tamaño de efecto es grande, por tanto, podemos rechazar la hipótesis nula H.1.2.0 y aceptar que existe diferencia significativa en la EFICACIA del usuario al incorporar el MU-ABR o al no incorporarlo. Sin embargo, podemos aceptar la hipótesis nula H.1.3.0: No existe diferencia significativa en la SATISFACCION del usuario al incorporar el MU-ABR o al no incorporarlo.

4.2.2 RQ2: Retroalimentación de Progreso (MU-PFB)

Para este mecanismo de usabilidad, los resultados son muy dispares dependiendo del atributo de calidad y de las medidas seleccionadas. Considerando la **Eficiencia**, los sujetos sin acceso al mecanismo parecen invertir más número de clics que los que tienen acceso al mismo. Este probable incremento en los clics del usuario se interpretaría como una pérdida temporal del control del sistema o una desorientación del usuario debido a la sensación de inactividad o retardo derivado de la tarea de búsqueda. Sin embargo, aunque existan diferencias estadísticamente significativas, el tamaño de efecto es medio ocasionando que esta diferencia en número de clics sea irrelevante (~4 clics). En cuanto al tiempo, las diferencias no son significativas y el tamaño de efecto es despreciable. No hay mejora en la rapidez del usuario. Creemos que como este mecanismo informa al usuario de lo que está sucediendo en el sistema mientras se ejecuta la tarea, es razonable pensar que los tiempos de ejecución de cada usuario tienda hacia los mismos valores, independientemente de la presencia o no del mecanismo. Concluimos que se acepta la hipótesis nula H.2.1.0: No existe diferencia significativa en la EFICIENCIA del usuario al incorporar el MU-PFB o al no incorporarlo.

En términos de **Eficacia**, las evidencias obtenidas suponen que la presencia o no del mecanismo no ha sido determinante para la finalización exitosa de la tarea. La diferencia no es estadísticamente significativa. El resultado no es tan sorprendente debido al retardo imperceptible del sistema. El poder de este mecanismo es hacer que las aplicaciones lentas sean más tolerables y que los usuarios sientan que se les está escuchando mediante una retroalimentación apropiada sobre sus acciones. Esto evita que los usuarios abandonen la tarea en curso ante esperas largas. Nuestra aplicación tiene respuestas relativamente rápidas, y por ello, los sujetos han completado la tarea con o sin el mecanismo presente. Concluimos que se acepta la hipótesis nula H.2.2.0: No existe diferencia significativa en la EFICACIA del usuario al incorporar el MU-PFB o al no incorporarlo.

Considerando la **Satisfacción**, las mejoras percibidas por los usuarios son despreciables por lo que aceptamos la hipótesis nula H.2.3.0: No existe diferencia significativa en la SATISFACCION del usuario al incorporar el MU-PFB o al no incorporarlo.

4.2.3 RQ3: Preferencias (MU-PRF)

Por último, en este mecanismo de usabilidad no se observan efectos significativos con respecto a la **Eficiencia**, ni en el tiempo ni en el número de clics (~5 clics y 31 seg). La diferencia no es significativa en ambos casos. Parece que la acción de este mecanismo en la personalización de la interfaz no ha ayudado a mejorar el desempeño del usuario debido a la sencillez de la tarea propuesta, el cual consiste en pasearse visualmente por la tienda buscando solamente información de plazos. Pensamos que podrían observarse efectos distintos en la realización de tareas más complejas entre una interfaz poco legible y una interfaz personalizable. Podemos aceptar la hipótesis nula H.3.1.0: No existe diferencia significativa en la EFICIENCIA del usuario al incorporar el MU-PRF o al no incorporarlo.

Considerando la **Eficacia**, además de la existencia de diferencias significativas, el tamaño de efecto medio proporciona un indicativo relevante de que los usuarios han logrado avanzar en mayor proporción de la tarea cuando el mecanismo está presente.

Para la **Satisfacción**, las diferencias significativas y el tamaño de efecto grande denotan que los usuarios valoraron positivamente la presencia de opciones configurables del sistema y que dichas configuraciones han facilitado la realización de la tarea. Concluimos que se rechaza las hipótesis nulas de Eficacia y Satisfacción (H.3.2.0 y H.3.3.0) y aceptamos que existe diferencia significativa en la EFICACIA Y SATISFACCIÓN del usuario al incorporar el MU-PRF o al no incorporarlo.

5. Agregación de Resultados

5.1 Enfoque de Análisis

Un total de tres experimentos constituyen la familia (es decir, el experimento original [Bouza, 2016] con un total de 100 sujetos, la primera réplica [Ferreira, 2017] compuesto de 168 sujetos y esta segunda réplica con 99 sujetos). Dividimos el análisis de estos tres conjuntos de 267 datos a nivel de familia en tres apartados diferentes, uno por cada mecanismo de usabilidad: Abortar Operación, Retroalimentación de Progreso y Preferencias. Se evalúan un total de cuatro variables respuesta por sección: CLICK (número de clics), ELAPSED_TIME (el tiempo transcurrido en la tarea), PERCETAGETASK (el porcentaje de la tarea completada) y VALUE (puntuación de satisfacción) en una escala Likert (de 1 a 5). En el experimento original [Bouza, 2016], no se midió la variable respuesta PERCETAGETASK.

estos tres conjuntos de 267 datos a nivel de familia por cada mecanismo de usabilidad.

Dentro de cada apartado (es decir, por mecanismo de usabilidad) seguimos un procedimiento idéntico:

- Proporcionamos un gráfico de perfil que muestra el puntaje promedio de los sujetos en los experimentos por cada variable respuesta y por la presencia/ausencia de los mecanismos de usabilidad. Las observaciones preliminares se hacen a nivel de familia con respecto a las diferencias de resultados de los experimentos.
- Siguiendo las convenciones usadas en medicina para analizar grupos de experimentos interrelacionados, se ajusta un modelo de regresión lineal de efectos fijos con el factor principal TRATAMIENTO y EXPERIMENTO para analizar los datos a nivel de familia [Debray et al., 2015; Whitehead, 2002]. Se eligió una regresión lineal sobre el meta-análisis de los tamaños del efecto [Borenstein et al., 2011] ya que: (1) el acceso a los datos brutos está garantizado dentro de la familia; y (2) todos los experimentos tienen operacionalizaciones de variable respuesta idénticas. Se seleccionó un modelo de regresión lineal de efectos fijos donde: (1) los experimentos son replicaciones exactas (es decir, repeticiones); (2) las poblaciones son similares en todos los experimentos; (3) se ha llevado a cabo un bajo número de experimentos. Se deben cumplir dos supuestos relevantes dentro de los modelos lineales de efectos fijos: el supuesto de normalidad y el supuesto de homocedasticidad (es decir, la igualdad de varianzas entre los grupos de tratamiento [Whitehead, 2002]). La suposición de normalidad es sostenible debido al tamaño de muestra relativamente grande logrado a nivel familiar (es decir, tamaño de muestra en cientos [Lumley et al., 2002; Vickers, 2005]). Los modelos de mínimos cuadrados generalizados [Pinheiro & Bates, 2000] que acomodan diferentes varianzas entre los grupos de tratamiento y los experimentos (es decir, que permiten la heterocedasticidad) se ajustaron con el objetivo de observar la solidez de los resultados de la regresión lineal. Como la regresión lineal y los resultados de mínimos cuadrados generalizados fueron similares, los de regresión

lineal (es decir, el modelo parsimonioso) finalmente se seleccionaron para interpretar la significación estadística de los resultados.

- La significación estadística de los resultados se juzga con base en el p-valor de la estimación de TRATAMIENTO a nivel familiar. La importancia práctica de los resultados se juzga en función de: (1) el signo de la estimación de TRATAMIENTO y (2) la magnitud de la estimación de TRATAMIENTO en comparación con el término Intercepto (es decir, la condición ausente de la primera réplica que se toma como la clase de referencia en todas las regresiones lineales). En caso de que las condiciones de control difieran marcadamente en los experimentos, la estimación de la condición de control de la primera réplica puede ser poco informativa para juzgar de la magnitud de estimación del TRATAMIENTO. La magnitud de la estimación del TRATAMIENTO se considera teniendo en cuenta también las estimaciones de control del resto de los experimentos analizados.

5.2 Análisis de Abortar Operación

La Figura 5 muestra el gráfico de perfil para CLICK, ELAPSED_TIME, PERCENTAGETASK y VALUE por presencia/ausencia del mecanismo de usabilidad de Abortar Operación en todos los experimentos.

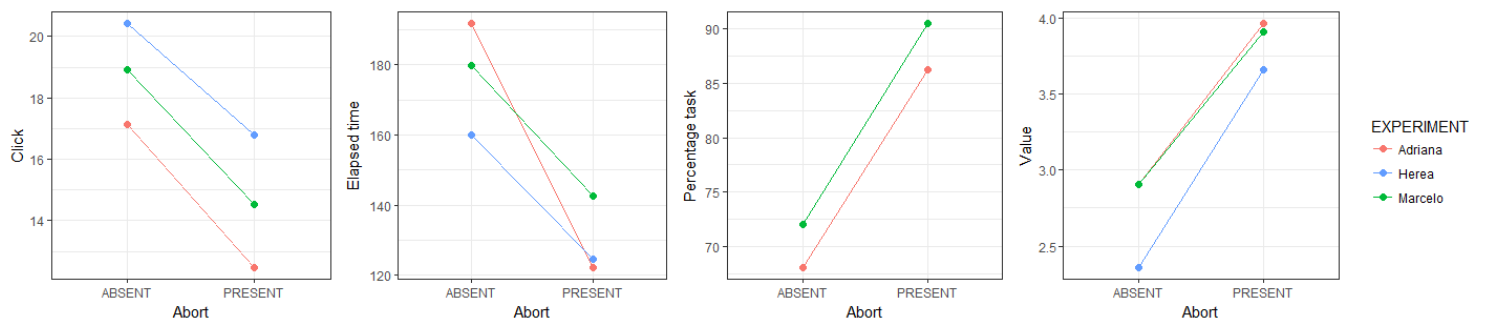


Figura 5 : Gráfico de perfil para Abortar Operación

Como se puede ver en la Figura 5, los promedios coinciden en gran medida en los experimentos: mientras que el mecanismo de cancelación disminuye el número de clics y el tiempo transcurrido en todos los experimentos, la presencia del mecanismo de Abortar Operación aumenta el porcentaje de finalización de la tarea y la satisfacción de los sujetos.

La Tabla 17 muestra los resultados de los modelos de regresión lineal ajustados de Abortar Operación en los experimentos.

Coeficiente	CLICK	ELAPSED_TIME	PERCENTAGETASK	VALUE
Intercepto	18.85 (1.21)***	183.76 (10.99)***	72.07 (2.05)***	2.86 (0.11)***
Presente	-4.25 (1.36)**	-45.16 (12.33)***	18.36 (2.56)***	1.09 (0.12)***
Experimento==Adriana	-1.93 (1.65)	-4.68 (14.99)	-4.07 (2.65)	0.03 (0.15)
Experimento==Herea	1.87 (1.64)	-18.90 (14.90)		-0.40 (0.15)**

Tabla 17: Coeficientes de regresión lineal para Abortar Operación

Niveles de Significancia: *** ($p < 0.001$), ** ($p < 0.01$), * ($p < 0.05$), ' ' ($p < 0.1$).

Como se puede ver en la Tabla 17, el mecanismo de usabilidad de Abortar Operación parece afectar en una medida media al número de clics (es decir, un descenso en el número de clics de alrededor del 20% con respecto a la intercepción, el puntaje promedio calculado para la condición de Abortar Operación ausente en la primera réplica), en mayor medida tanto en el tiempo transcurrido (es decir, una caída en el tiempo de alrededor del 25%) como en el porcentaje de finalización de tareas (es decir, una caída en el tiempo de alrededor del 25%) y en un grado mayor a la satisfacción (es decir, un aumento de casi el 40%). Por lo tanto, **la presencia o ausencia del mecanismo de Abortar Operación parece afectar en gran medida a la usabilidad del sistema.**

5.3 Análisis de Retroalimentación de Progreso

La Figura 6 muestra la gráfica de perfil para CLICK, ELAPSED_TIME, PERCETANGETASK y VALUE por presencia/ausencia del mecanismo de usabilidad de Retroalimentación de Progreso en todos los experimentos.

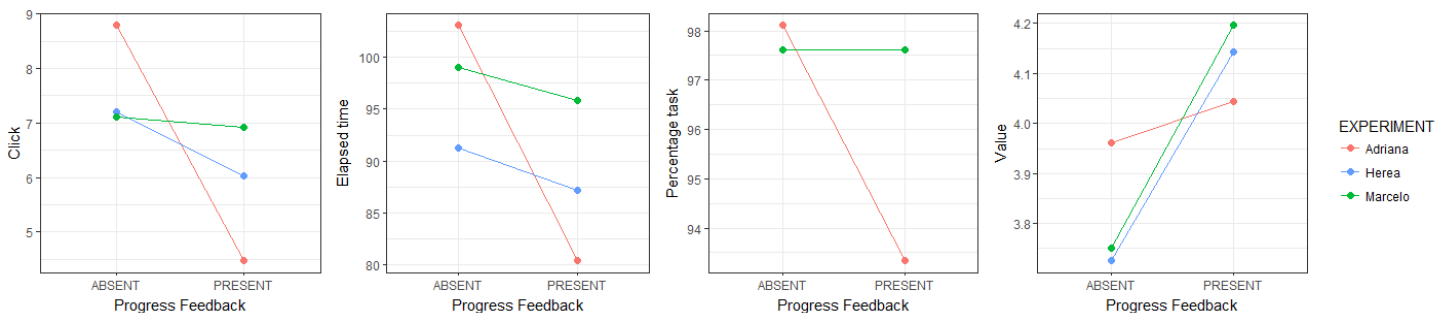


Figura 6: Gráfico de perfil para Retroalimentación de Progreso

Como se puede ver en la Figura 6, los sujetos de esta segunda réplica parecen experimentar una caída mucho mayor en la cantidad media de clics y el tiempo transcurrido que los del experimento original y la primera réplica. Además, los sujetos de esta segunda réplica obtienen una caída en el porcentaje de completitud que los de la primera réplica (a pesar de que la pendiente parece sustancial, la diferencia entre el punto más alto y el punto más bajo es solo un 4%). Finalmente, relativamente a los datos del experimento original y la primera réplica, los sujetos de esta segunda réplica obtienen un aumento menor en la satisfacción. Puede haber un efecto de interacción entre experimentos. Tal interacción puede ser objeto de un análisis posterior (por ejemplo, buscando variables moderadoras).

La Tabla 18 muestra los resultados de los modelos de regresión lineal ajustados para analizar la presencia o ausencia del mecanismo de Retroalimentación de Progreso en los experimentos.

Coefficiente	CLICK	ELAPSED_TIME	PERCENTAGETASK	VALUE
Intercepto	7.73 (0.61)***	101.02 (7.74)***	98.50 (1.69)***	3.79 (0.08)***
Presente	-1.43 (0.60)*	-7.18 (7.58)	-1.75 (2.10)	0.37 (0.08)***
Experimento==Adriana	-0.27 (0.88)	-5.07 (11.13)	-1.77 (2.18)	0.04 (0.12)
Experimento==Herea	-0.39 (0.68)	-8.20 (8.61)		-0.04 (0.09)

Tabla 18: Coeficientes de regresión lineal para Retroalimentación de Progreso

Niveles de Significancia: *** ($p < 0.001$), ** ($p < 0.01$), * ($p < 0.05$), ‘.’ ($p < 0.1$).

Como se puede ver en la Tabla 18, **el mecanismo de usabilidad de Retroalimentación de Progreso parece afectar en menor medida a las variables respuesta que el resto de los mecanismos de usabilidad**. Específicamente, el número de clics y el tiempo transcurrido se reducen en aproximadamente un 20% y 7% respectivamente, el porcentaje de finalización de tareas es insignificante (alrededor del 1%). Además, la satisfacción parece no aumentar mucho con respecto a la condición de control (solo 10%). En vista de estos hallazgos, **la presencia o ausencia del mecanismo de Retroalimentación de Progreso parece no afectar en gran medida a la usabilidad del sistema**. Pueden existir interacciones y, por lo tanto, los moderadores deben identificarse con el objetivo de explicar la heterogeneidad de los resultados.

5.4 Análisis de Preferencias

La Figura 7 muestra el gráfico de perfil para CLICK, ELAPSED_TIME, PERCENTAGETASK y VALUE por presencia/ausencia del mecanismo de usabilidad de Preferencias en todos los experimentos.

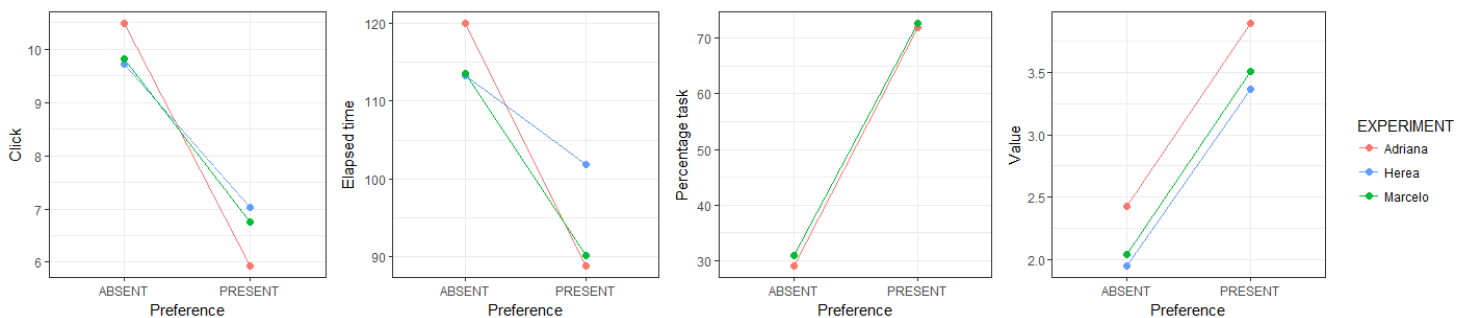


Figura 7: Gráfico de perfil para Preferencias

Como se puede ver en la Figura 7, las medias coinciden en gran medida en los experimentos: mientras que el mecanismo de Preferencias disminuye el número de clics y el tiempo transcurrido en todos los experimentos, la presencia del mecanismo aumenta el porcentaje de finalización de tareas y la satisfacción de los sujetos con un grado resaltable.

La Tabla 19 muestra los resultados de los modelos de regresión lineal ajustados para analizar la presencia o ausencia del mecanismo de Preferencias a través de los experimentos.

Coeficiente	CLICK	ELAPSED_TIME	PERCENTAGETASK	VALUE
<i>Intercept</i>	9.86 (1.10)***	111.29 (8.80)***	30.74 (4.49)***	2.05 (0.11)***
<i>Presente</i>	-3.15 (1.08)**	-18.79 (8.61)*	42.10 (5.59)***	1.44 (0.11)***
<i>Experimento==Adriana</i>	-0.09 (1.58)	2.39 (12.64)	-1.19 (5.79)	0.38 (0.15)*
<i>Experimento==Herea</i>	0.09 (1.23)	5.86 (9.79)		-0.11 (0.12)

Tabla 19: Coeficientes de regresión lineal para Preferencias

Niveles de Significancia: *** (p<0.001), ** (p<0.01), * (p<0.05), ‘.’ (p<0.1).

Como se puede ver en la Tabla 19, el mecanismo de usabilidad de Preferencias parece afectar en gran medida al número de clics (es decir, una caída en el número de clics alrededor del 33%), en una pequeña medida al tiempo transcurrido (es decir, una disminución del tiempo de aproximadamente el 15%), en gran medida hasta la finalización de la tarea (es decir, un aumento de casi el 133%) y en gran medida a la satisfacción (un aumento de alrededor del 75%). Por lo tanto, **la presencia o ausencia del mecanismo de usabilidad de Preferencias parece afectar sustancialmente a la usabilidad del sistema.**

6. Conclusiones y Trabajo Futuro

6.1 Conclusión

Hemos estudiado empíricamente los mecanismos de usabilidad en una aplicación web con el propósito de evaluar el impacto de usabilidad desde la perspectiva de los usuarios en el contexto de usuarios no informáticos. Los mecanismos sometidos al estudio empírico son: Abortar Operación, Retroalimentación de Progreso y Preferencias. La evaluación de cada mecanismo se realiza respecto a tres atributos de calidad: eficiencia, eficacia y satisfacción.

Para ello, hemos llevado a cabo un experimento en el cual se definieron los distintos componentes necesarios para el estudio empírico. La aplicación y su interfaz están diseñadas para la asignación automática de grupos y tareas así como la recolección de datos. De esta manera, hemos logrado comparar si la usabilidad percibida por el grupo que ha interactuado con la aplicación teniendo el mecanismo presente era mejor o no que el grupo que no lo ha tenido.

A nivel global, las evidencias obtenidas tienden a favor de una mejora en la usabilidad del sistema. No hemos obtenido evidencias de que la usabilidad vaya en detrimento de la interacción del usuario con el sistema. En tal sentido, los resultados sugieren lo siguiente:

- En el caso de la variable respuesta eficiencia los resultados son reveladores. Aunque en ningún caso la existencia del mecanismo de usabilidad perjudica la eficiencia de los usuarios, no siempre se observa una mejora determinante. En el caso del mecanismo Abortar Operación, el tamaño de efecto es pequeño por lo que no parece ser relevante y entendemos que la mejora en la eficiencia del usuario es mínima. En el caso del mecanismo Retroalimentación de Progreso el tamaño de efecto es medio en el caso del número de clics por lo que se puede apreciar una mejoría destacable, sin embargo, en el caso de la rapidez la mejoría no es notable. Por último para el mecanismo Preferencias, la diferencia entre la presencia o ausencia del mecanismo es irrelevante en términos de eficiencia.
- En el caso de la variable respuesta eficacia para el mecanismo Abortar Operación la diferencia entre la presencia y ausencia del mecanismo es grande, por lo que la mejora es concluyente. En el mecanismo Retroalimentación de Progreso no se aprecia una mejora significativa. Por último en el mecanismo Preferencias la diferencia es nuevamente destacable por lo que hay una mejoría clara.
- Para la variable respuesta satisfacción, tanto en el caso de Abortar Operación como en el mecanismo Preferencias existe diferencia significativa entre la ausencia y presencia del mecanismo, mientras que en el caso de Retroalimentación de Progreso la mejoría es prácticamente inexistente.

La segunda parte del estudio consiste en un meta-análisis aplicado a un conjunto de tres experimentos (el experimento original, la primera réplica y esta segunda réplica). Para ello, hemos utilizado el modelo de regresión lineal aplicado a nivel de familia y teniendo como referencia la primera réplica.

Así, hemos concluido que la presencia o ausencia del mecanismo Abortar Operación parece afectar en gran medida a la usabilidad del sistema. El mecanismo de usabilidad Retroalimentación de Progreso parece afectar en menor medida a las variables respuesta que el resto de mecanismos de usabilidad, mientras que la presencia o ausencia del mecanismo de usabilidad Preferencias parece afectar sustancialmente a la usabilidad del sistema.

6.2 Trabajo Futuro

Esta investigación es un paso para el análisis empírico del impacto de la usabilidad desde la perspectiva del usuario. Además, la existencia de diferencias estadísticamente significativas pero con tamaño de efecto pequeño proporciona razón suficiente para seguir la experimentación y las siguientes líneas resultan prometedoras:

- Llevar a cabo más réplicas del experimento con sujetos distintos a la muestra descrita en esta investigación.
- Estudiar el impacto de otros mecanismos de usabilidad desde la perspectiva de los usuarios en los mismos atributos de calidad.
- Estudiar el impacto de los mecanismos de usabilidad sobre otros atributos de calidad.
- Replantear el diseño experimental en términos de la instrumentación para aplicarlo a otro dominio que no sea una tienda en línea.
- Replantear el diseño de las tareas para la creación de escenarios más complejos.

Finalizando, los estudios empíricos van tomando cada vez más importancia en IS generando evidencias que sustentan la evolución del conocimiento a fin de ir mejorando las teorías como consecuencia de los resultados empíricos.

Referencias

S. T. Acuña, M. Gómez and N. Juristo, "Towards Understanding the Relationship between Team and Software Quality-A Quasi-Experimental Study", *Empirical Software Engineering*, vol.13, no. 4, pp. 401-434, 2008.

M. Alshamari and P. Mayhew, "Technical Review: Current Issues of Usability Testing", *IETE Technical Review*, vol. 26, no. 6, p. 402, 2009.

H. Amatriain, E. Diez, R. Bertone and E. Fernández, "Establecimiento del Modelo de Agregación más Adecuado para Ingeniería del Software", *Revista Latinoamericana de Ingeniería de Software*, vol. 2, no. 6, p. 369, 2014.

M. Aveledo, D. M. Curtino, A. De la Rosa A, A. M. Moreno, "Measuring the Effect of Usability Mechanisms On User Efficiency, Effectiveness and Satisfaction", In: *Proceedings of the 24th International Conference of Software Engineering and Knowledge Engineering SEKE*, pp. 599–604, 2012.

J. Battey, "IBM's Redesign Results in a Kinder", *Simpler Web Site*, 2001. Available at: http://interface.free.fr/Archives/IBM_redesign_results.pdf

J. Black, *Usability Is Next to Profitability*, 2002. Available at: http://www.businessweek.com/technology/content/dec2002/tc2002124_2181.htm

D. Bolchini, F. Garzotto and F. Sorce, "Does Branding Need Web Usability? A Value-Oriented Empirical Study", In: *Human-Computer Interaction -- INTERACT 2009: 12th IFIP TC 13 International Conference*. Berlin: Springer, pp 652–665, 2009.

M. Borenstein, L. V. Hedges, J. P. Higgins and H. R. Rothstein, *Introduction to meta-analysis*. Chichester, U.K.: John Wiley & Sons, 2011.

H. Bouza, "Evaluación Empírica de los Mecanismos de Usabilidad de Preferencias, Retroalimentación de Progreso y Abortar Operación en un Entorno Web", *Trabajo de Fin de Grado*. Directora: Silvia T. Acuña. Doble Grado en Ingeniería Informática-Matemáticas, Escuela Politécnica Superior, Universidad Autónoma de Madrid, 2016.

W. C. Cochran and G. M. Cox, *Diseños Experimentales*. Distrito Federal, México: Trillas, 1980.

V. de Castro, M. Genero, E. Marcos and M. Piattini, "Empirical Study to Assess whether the Use of Routes Facilitates the Navigability of Web Information Systems", *IET Software*, vol. 5, no. 6, p. 528, 2011.

T. Debray, K. Moons, G. van Valkenhoef, O. Efthimiou, N. Hummel, R. Groenwold and J. Reitsma, "Get Real in Individual Participant Data (IPD) Meta-Analysis: A Review of the Methodology", *Research Synthesis Methods*, vol. 6, no. 4, pp. 293-309, 2015.

G. M. Donahue, "Usability and the Bottom Line", *IEEE Software*, vol. 18, no. 1, pp. 31-37, 2001.

D. Falessi, N. Juristo, C. Wohlin, B. Turhan, J. Münch, A. Jedlitschka and M. Oivo, "Empirical Software Engineering Experts on the Use of Students and Professionals in Experiments", *Empirical Software Engineering*, vol. 23, no. 1, pp. 452-489, 2017.

X. Fang and C. Holsapple, "An Empirical Study of Web Site Navigation Structures' Impacts on Web Site Usability", *Decision Support Systems*, vol. 43, no. 2, pp. 476-491, 2007.

J. M. Ferreira, "Impacto de los Mecanismos de Usabilidad sobre la Eficiencia, Eficacia y Satisfacción del Usuario: Una Evaluación Empírica", Tesis de Máster. Directora: Silvia T. Acuña. Magíster en Tecnologías de la Información y Comunicación, Facultad de Informática, Universidad Nacional de Asunción, 2017.

J. M. Ferreira and S. T. Acuña, "A Software Application for Collecting Usability Empirical Data about User Efficiency, Effectiveness and Satisfaction", In: *Proceedings of the XII Iberoamerican Conference on Software Engineering and Knowledge Engineering (JIISIC'2017)*, p. 11, 2017.

A. Field, J. Miles and Z. Field, "Discovering Statistics Using R". London, United Kingdom: SAGE, 2012.

Y. Hassan Montero and F. Martín Fernández, "A Proposal of User-Centered Design Methodology for the Development of Accessible Web Sites", *Revista Española de Documentación Científica*, vol.27, no. 3, 2004.

ISO 9241-11. Ergonomics Requirements for Office Work with Visual Display Terminals (VDTs) - Part 11: Guidance on Usability, 1998.

M. Ivory and M. Hearst, "The State of the Art in Automating Usability Evaluation of User Interfaces", *ACM Computing Surveys*, vol. 33, no. 4, pp. 470-516, 2001.

N. Juristo and A. M. Moreno, "Basics of Software Engineering Experimentation". NY: Springer, 2010.

N. Juristo, A. Moreno and M. Sánchez-Segura, "Analysing the Impact of Usability on Software Design", *Journal of System and Software*, vol. 80, no. 9, pp. 1506-1516, 2007a.

N. Juristo, A. Moreno and M. Sánchez-Segura, "Guidelines for Eliciting Usability Functionalities", *IEEE Transactions on Software Engineering*, vol. 33, no. 11, pp. 744-758, 2007b.

A. Ko, T. D. LaToza and M. M. Burnett, "A Practical Guide to Controlled Experiments of Software Engineering Tools with Human Participants", *Empirical Software Engineering*, vol. 20, no. 1, pp. 110-141, 2013.

T. Lumley, P. Diehr, S. Emerson and L. Chen, "The Importance of the Normality Assumption in Large Public Health Data Sets", *Annual Review of Public Health*, vol. 23, no. 1, pp. 151-169, 2002.

- X. Metzger, “La Agregación de Datos en la Medición de Desigualdades e Inequidades en la Salud de las Poblaciones”, *Revista Panamericana de Salud Pública*, vol. 12, no. 6, pp. 445-453, 2002.
- J. Nielsen, *Usability Engineering*. Boston: Academic Press, 1993.
- J. Panach, N. Juristo and O. Pastor, “Including Functional Usability Features in a Model-driven Development Method”, *Computer Science and Information Systems*, vol. 10, no. 3, pp. 999-1024, 2013.
- L. Perurena y M. Moráquez, “Usabilidad de los Sitios Web, los Métodos y las Técnicas”, *Revista Cubana de Información en Ciencias de la Salud*, vol. 24, no. 2, pp. 176-194, 2013.
- J. Pinheiro and D. Bates, *Mixed-effects models in S and S-PLUS*. New York: Springer, 2000.
- QuickStore (Experimental Application), 2016. Available at: <http://webadm.senado.gov.py/tesisweb/>
- W. Sánchez, “La Usabilidad en Ingeniería de Software: Definición y Características”, *Innovación. Revista de Ingeniería e Innovación de la Facultad de Ingeniería Don Bosco*, Madrid, no. 2, pp. 7-21, 2011.
- J. Sauro and E. Kindlund, “A Method to Standardize Usability Metrics Into a Single Score”, In: *Proceedings of the SIGHI Conference on Human Computing System*, pp. 401-409, ACM, 2005.
- D. Sjøeberg, J. Hannay, O. Hansen, V. Kampenes, A. Karahasanovic, N. Liborg and A. Rekdal, “A Survey of Controlled Experiments in Software Engineering”, *IEEE Transactions on Software Engineering*, vol. 31, no. 9, pp. 733-753, 2005.
- M. van Welie, “The Amsterdam Collection of Patterns in User Interface Design”, 2008. Available at <http://www.welie.com/patterns/>
- A. Vickers, “Parametric versus Non-Parametric Statistics in the Analysis of Randomized Trials with Non-Normally Distributed Data”, *BMC Medical Research Methodology*, vol. 5, no. 1, 2005.
- A. Whitehead, *Meta-analysis of controlled clinical trials*. U.K.: John Wiley & Sons, vol. 7, 2002.
- J. Winter and K. Rönkkö, “SPI Success Factors within Product Usability Evaluation”, *Journal of Systems and Software*, vol. 83, pp. 2059-2072, 2010.
- C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell and A. Wesslén, *Experimentation in Software Engineering: An Introduction*. Berlin: Springer, 2012.

Glosario

Asimetría: en estadística, las medidas de asimetría son indicadores que permiten establecer el grado de simetría (o asimetría) que presenta una distribución de probabilidad de una variable aleatoria sin necesidad de realizar su representación gráfica. En este estudio se utiliza el *Coeficiente de Asimetría de Fisher* que utiliza el tercer momento estándar y es la medida de asimetría más utilizada.

Curtosis: en teoría de la probabilidad y estadística, las medidas de curtosis tratan de estudiar la proporción de la varianza que se explica por la combinación de datos extremos respecto a la media en contraposición con datos poco alejados de la forma.

Estudio empírico: modelo de investigación científica que se basa en la experimentación y la lógica empírica que junto con la observación de fenómenos y su análisis estadístico, es el más usado en el campo de las ciencias sociales y las ciencias naturales.

Evidencia empírica: información que es adquirida a través de la observación o de la experimentación.

Experimento verdadero: diseños de investigación experimentales en los cuales los sujetos o grupos de sujetos de estudio están asignados aleatoriamente a los tratamientos.

Réplica: es una repetición, en nuestro caso experimental, con la misma configuración de factores. Las réplicas están sujetas a las mismas fuentes de variabilidad, de forma independiente unas de otras.

Agregación: la síntesis cuantitativa consiste en integrar los resultados de un conjunto de experimentos en una medida resumen. Con esto se busca hallar un resultado que sea resumen representativo de los resultados de los estudios individuales y que signifique una mejora sobre las estimaciones individuales. Este procedimiento es la agregación.

Meta-análisis: es el análisis estadístico de una colección de resultados procedentes de estudios individuales y cuyo propósito es integrar dichos resultados. Supone una alternativa rigurosa frente a las discusiones meramente narrativas sobre los resultados de una colección de estudios.

Delta de Cliff: es un estadístico que permite cuantificar la magnitud de la diferencia entre dos grupos de observaciones que resultan incompatibles con el supuesto de normalidad. El análisis de esta cuantificación complementa la interpretación del p-valor asociado a la correspondiente prueba de hipótesis.

U de Mann-Whitney: en estadística es una prueba no-paramétrica aplicada a dos muestras independientes. Se usa para comprobar la heterogeneidad de dos muestras ordinales.

Modelo de efectos fijos: modelo estadístico que representa las cantidades observadas en las variables explicativas que son tratadas como si las cantidades fueran no-aleatorias. Esto está en contraste con el **Modelo de efectos aleatorios** y el **Modelo mixto** en los que todas o algunas de las variables explicativas son tratadas como si se derivaran de causas aleatorias.

Anexos

A. Cuestionario de Familiaridad

En este Anexo se detalla el cuestionario de familiaridad presentado en el experimento verdadero, éste se recoge en la Figura 5.

Nombre o Alias *	<input type="text"/>
Datos Opcionales:	(Correo Electrónico Opcional) <input type="text"/> (Escribe tu correo electrónico) (Nro. de Teléfono Opcional) <input type="text"/> (Escribe tu nro. de teléfono)
Sexo: *	(Una selección) <input type="radio"/> MASCULINO <input type="radio"/> FEMENINO
Edad *	(Una selección) <input type="radio"/> MENOS DE 18 <input type="radio"/> 18 - 30 <input type="radio"/> 31 - 40 <input type="radio"/> 41 - 50 <input type="radio"/> 55 O MAS
Idiomas conocidos *	(Una o más selecciones) <input type="checkbox"/> CASTELLANO <input type="checkbox"/> INGLÉS <input type="checkbox"/> FRANCÉS <input type="checkbox"/> OTROS
¿Eres profesional en el Área de Informática/Computación? *	(Una selección) <input type="radio"/> SI <input type="radio"/> NO
Profesión/Ocupación:	<input type="text"/>
¿Dónde usas Internet con más frecuencia? *	(Una selección con opción de texto adicional) <input type="radio"/> EN CASA <input type="radio"/> EN EL TRABAJO <input type="radio"/> NO USO INTERNET <input type="radio"/> OTROS <input type="text"/> Otros usas
¿Para qué usas Internet? *	(Una o más selecciones con opción de texto adicional) <input type="checkbox"/> COMPRAS <input type="checkbox"/> TRABAJO <input type="checkbox"/> EDUCACION <input type="checkbox"/> ENTRETENIMIENTO <input type="text"/> Otros
¿Qué aplicaciones web usas con mayor frecuencia? *	(Una o más selecciones con opción de texto adicional) <input type="checkbox"/> FACEBOOK <input type="checkbox"/> TWITTER <input type="checkbox"/> GOOGLE <input type="checkbox"/> YAHOO <input type="text"/> Otros
¿Has realizado compras personalmente por Internet? *	(Una selección) <input type="radio"/> NUNCA <input type="radio"/> RARA VEZ <input type="radio"/> A VECES <input type="radio"/> CASI SIEMPRE <input type="radio"/> SIEMPRE
¿Has realizado compras por Internet con ayuda de intermediarios? *	(Una selección) <input type="radio"/> NUNCA <input type="radio"/> RARA VEZ <input type="radio"/> A VECES <input type="radio"/> CASI SIEMPRE <input type="radio"/> SIEMPRE
¿Qué tiendas en Internet has usado para buscar o comprar artículos? *	(Una selección) <input type="checkbox"/> AMAZON <input type="checkbox"/> eBAY <input type="checkbox"/> OTRAS TIENDAS <input type="checkbox"/> NINGUNO
¿Qué problema percibes cuando compras o quieres comprar por Internet? *	(Una selección con opción de texto adicional) <div> <input type="radio"/> Los artículos se presentan con precios inesperados (impuestos, tasas, envíos, etc.) <input type="radio"/> Las tiendas físicas ofrecen mejores precios y promociones <input type="radio"/> Siento inseguridad en el pago (tarjetas de crédito, etc.) <input type="radio"/> Tengo dudas sobre el soporte y la garantía de los artículos <input type="radio"/> No encuentro problemas <input type="radio"/> OTROS </div> <input type="text"/> Otros problemas
En líneas generales, ¿consideras una opción atractiva las compras por Internet? *	(Una selección) <input type="radio"/> SI <input type="radio"/> NO

Figura 8: Cuestionario de Familiaridad de la aplicación Quickstore

B. Diseño de Tareas y Cuestionarios de Satisfacción Post-Tarea

En este Anexo se muestra el diseño de las tareas a realizar en el experimento así como sus respectivos cuestionarios de satisfacción post-tarea.

a)

*Imagina que en el día de ayer visitaste la tienda y ya tienes ítems o artículos agregados a tu carrito. Accede entonces a tu carrito de compra y de la lista de artículos ya agregados, **modifica la cantidad** del primer ítem/artículo ingresando la cantidad que desees, luego **ingresa el código promocional "2015"** en la casilla especificada y valida dicho código para tener el precio final. Ahora te das cuenta de que el precio final supera tus expectativas y te arrepientes, por lo que decides comprar en otro momento y **anulas las acciones para dejar sin efecto los cambios que acabas de realizar en tu carrito.***

Para hacer esta tarea accede a la Tienda Virtual haciendo click

b)

Sentencias	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo	Comentarios
1. He logrado hacer la cancelación	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
2. Me sentí cómodo haciendo la cancelación de las modificaciones que no deseo en mi carrito de compra	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
3. Siento que el sistema me ayuda porque me provee alerta de los cambios hechos en mi carrito de compra y me solicita confirmación para proceder	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>

Figura 9: Tarea propuesta y Cuestionrio Post-Tarea para el MU-ABR

a)

*Imagina que un amigo te ha recomendado visitar la tienda porque estás interesado en adquirir el libro titulado "LA CHICA DEL TREN". Entonces realiza una **búsqueda de este artículo** y agrégalo a tu carrito de compra.*

Para hacer esta tarea accede a la Tienda Virtual haciendo click

b)

Sentencias	Totalmente de acuerdo	De acuerdo	Neutral	En desacuerdo	Totalmente en desacuerdo	Comentarios
1. He logrado buscar el artículo deseado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
2. He tenido algún tipo de retroalimentación mientras estaba realizando la búsqueda. Es decir, en todo momento el sistema me informaba de lo que estaba haciendo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
3. Me he sentido orientado/ayudado por el sistema cuando hice la búsqueda del artículo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>

[Siguiente](#)

Figura 10: Tarea propuesta y Cuestionario Post-Tarea para el MU-PFB

a)

Revisa la apariencia de la Tienda Virtual y comprueba que estás cómodo con sus características: tamaño de letra, tipo de letra... **Si necesitas cambiar algo, ahora es el momento. Procede, por favor, a hacerlo.**

Para hacer esta tarea accede a la Tienda Virtual haciendo click [Hacer esta Tarea](#)

Imagina ahora que estás interesado en devolver un artículo adquirido. Para ello, necesitas saber el **plazo de tiempo** que tienes para hacer la devolución. Entonces pásate por la tienda para buscar algún acceso o enlace que te proporcione dicha información.

Para hacer esta tarea accede a la Tienda Virtual haciendo click [Hacer esta Tarea](#)

b)

Sentencias	Totalmen de acuerdo	De acuerdo	Neutral	En desacuer	Totalmen en desacuer	Comentarios
1. He logrado poner la interfaz del sistema de acuerdo a mis preferencias (tamaño y tipo de letra)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
2. He logrado encontrar la información de los plazos para la devolución de artículos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
3. Pude encontrar, de una manera sencilla , la información del plazo para la devolución con las configuraciones visuales de la tienda	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>
4. Las configuraciones visuales de la tienda me han ayudado a localizar más rápidamente la información del plazo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="text"/>

[Siguiente](#)

Figura 11: Tarea, subtarea y Cuestionario Post-Tarea para el MU-PRF

C. Diagramas de cajas

En este Anexo se muestran los Diagramas de caja de cada Mecanismo de Usabilidad que han sido evaluados durante el experimento (Abortar Operación, Retroalimentación de Progreso y Preferencias). Dentro de cada Mecanismo de Usabilidad se muestra un diagrama por cada variable respuesta.

C.1 Diagrama de cajas del Mecanismo de Usabilidad Abortar Operación

El *boxplot* que podemos observar en la Figura 12 muestra la comparación del número de clics de aquellos sujetos que poseían el mecanismo frente a aquellos que no los tenían. Ambas cajas presentan una apariencia similar, estando la caja del mecanismo ausente más desplazada hacia arriba. Con ello entendemos que los sujetos con el mecanismo de usabilidad ausente, parecen realizar más clics que aquellos con el mecanismo presente. Por último, observamos datos atípicos en ambos casos, concretamente un dato atípico por cada caja.

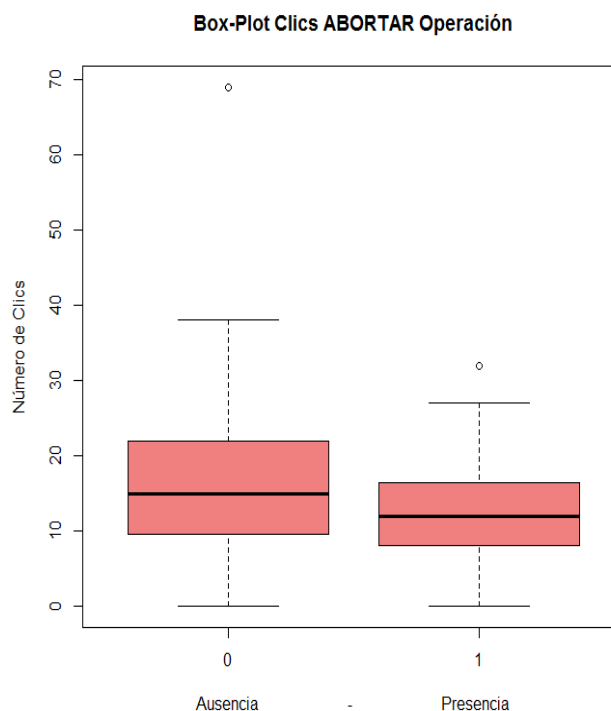


Figura 12: Diagrama de cajas para la variable N° de Clics en el Mecanismo de Usabilidad Abortar Operación

En el *boxplot* (Figura 13) podemos observar como la caja cuando el mecanismo está ausente está nuevamente desplazada hacia arriba. Esto nos revela que los sujetos con el mecanismo presente parece que utilizan menos tiempo para realizar la tarea que aquellos con el mecanismo ausente. Por último, observamos cuatro datos atípicos en el caso de la presencia.

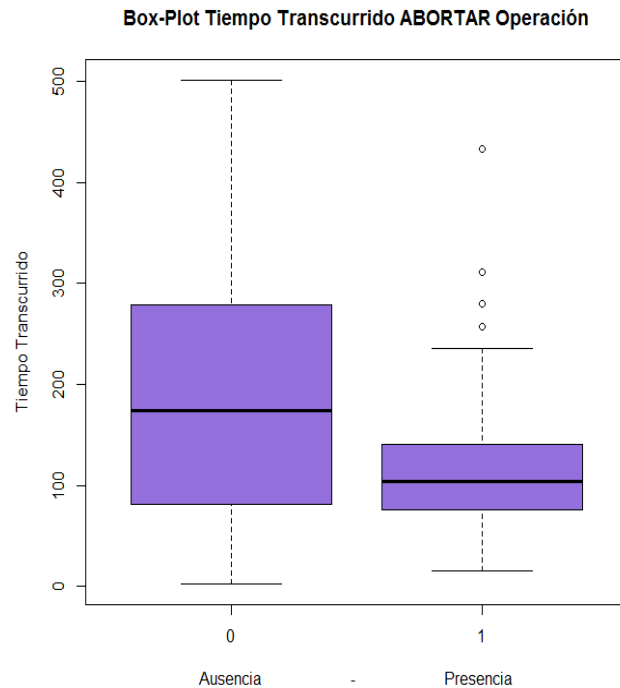


Figura 13: Diagrama de cajas para la variable Tiempo en el Mecanismo de Usabilidad Abortar Operación

En la Figura 14 se visualiza el *boxplot* del porcentaje de tarea realizada por cada sujeto, con respecto a la presencia y ausencia del mecanismo. En ella podemos observar como en el caso de presencia no hay bigote superior por lo que éste coincide con el tercer cuartil y la caja está desplazada para arriba dentro del rango de los valores superiores. En el caso de ausencia, carecemos de caja, únicamente tenemos la línea que nos muestra el valor de la mediana en 75 y dos datos atípicos al igual que en el caso de presencia.

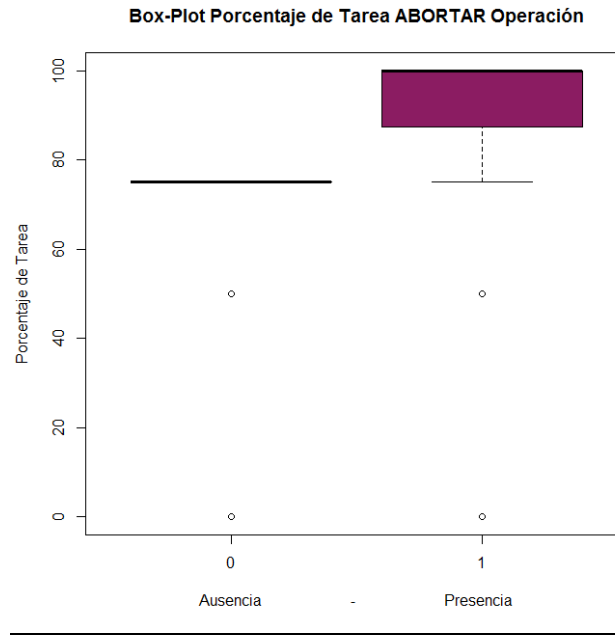


Figura 14: Diagrama de cajas para la variable Porcentaje de Tarea en el Mecanismo de Usabilidad Abortar Operación

Se puede apreciar en el *boxplot* (Figura 15) que las muestras difieren bastante y no reflejan simetría alguna. También, la gráfica del grupo que ha tenido el mecanismo presente refleja un sesgo a la izquierda sin bigote superior, es decir, el tercer cuartil coincide con el valor máximo. En este caso se observa únicamente un valor atípico correspondiente a la muestra que ha tenido el mecanismo presente.

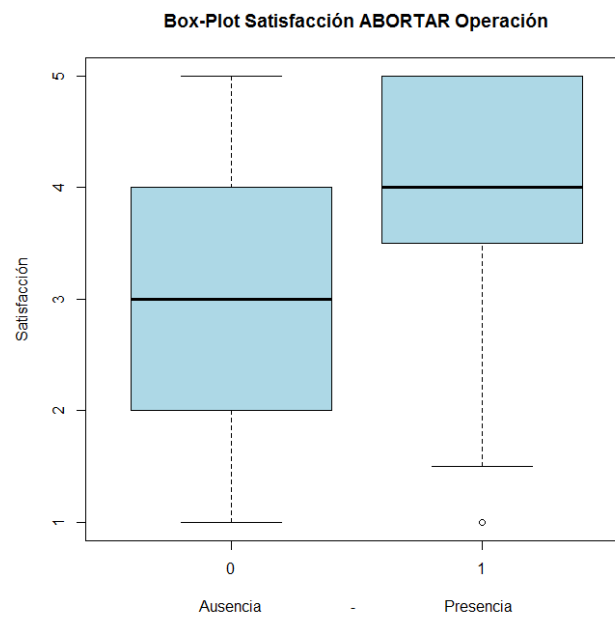


Figura 15: Diagrama de cajas para la variable Satisfacción en el Mecanismo de Usabilidad Abortar Operación

C.2 Diagrama de cajas del Mecanismo de Usabilidad Retroalimentación de Progreso

Si analizamos la Figura 16 vemos como las cajas son parecidas, estando más desplazada hacia abajo en el caso de la presencia. Esto es un indicador de que los sujetos con el mecanismo presente parecen utilizar menos clics que aquellos con el mecanismo ausente. También observamos varios valores atípicos, en concreto, tres en el caso de la ausencia y dos en el de la presencia.

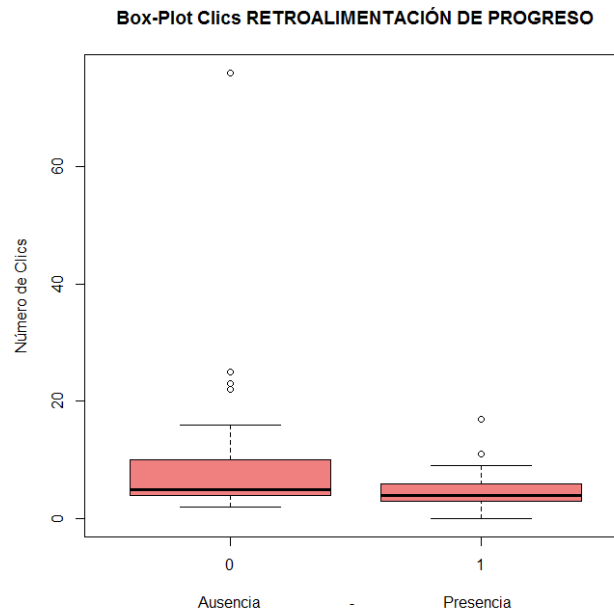


Figura 16: Diagrama de cajas para la variable N° de Clics en el Mecanismo de Usabilidad Retroalimentación de Progreso

Observando la Figura 17 podemos ver como ambas cajas son similares llegando a valores más altos la correspondiente con la ausencia del mecanismo. Esto nos muestra que los sujetos con el mecanismo ausente parecen utilizar más tiempo en resolver la tarea. Hay varios puntos atípicos, en concreto, dos en el caso ausente y tres en el presente.

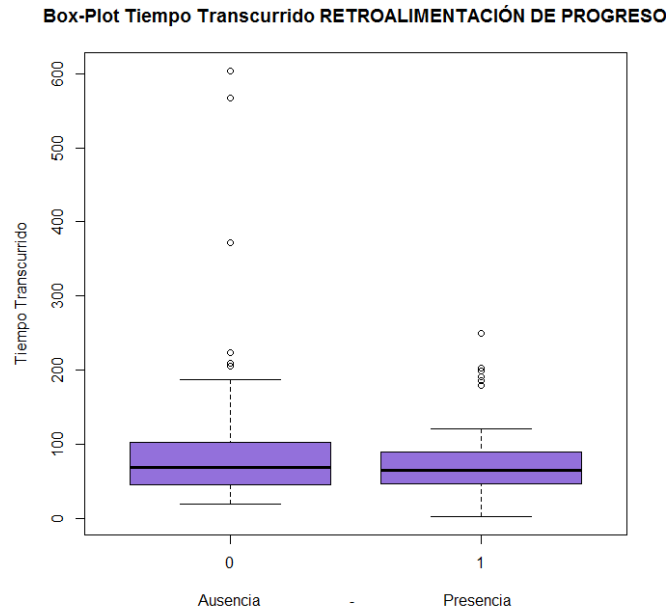


Figura 17: Diagrama de cajas para la variable Tiempo en el Mecanismo de Usabilidad Retroalimentación de Progreso

Observando la Figura 15, podemos ver como únicamente está marcada la línea correspondiente a la mediana y dos datos atípicos, uno en el mecanismo presente y otro en el ausente. Esto nos indica que prácticamente todos los sujetos finalizaron completamente la tarea.

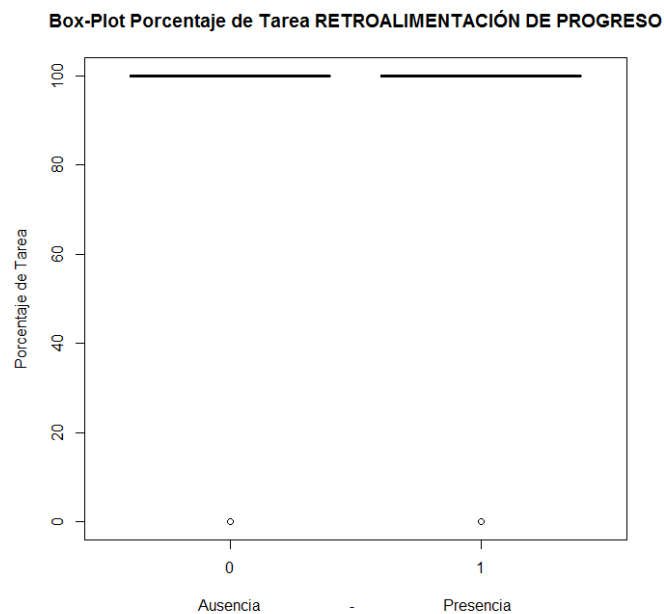


Figura 18: Diagrama de cajas para la variable Porcentaje de Tarea en el Mecanismo de Usabilidad Retroalimentación de Progreso

La Figura 19 nos muestra dos cajas principalmente desplazadas hacia arriba lo cual nos indica que los sujetos están en su mayoría satisfechos. Sin embargo, en el caso de la

ausencia tiende a desplazarse para abajo, tocando también los valores bajos de la gráfica. Se presenta un valor atípico en el caso de la presencia.

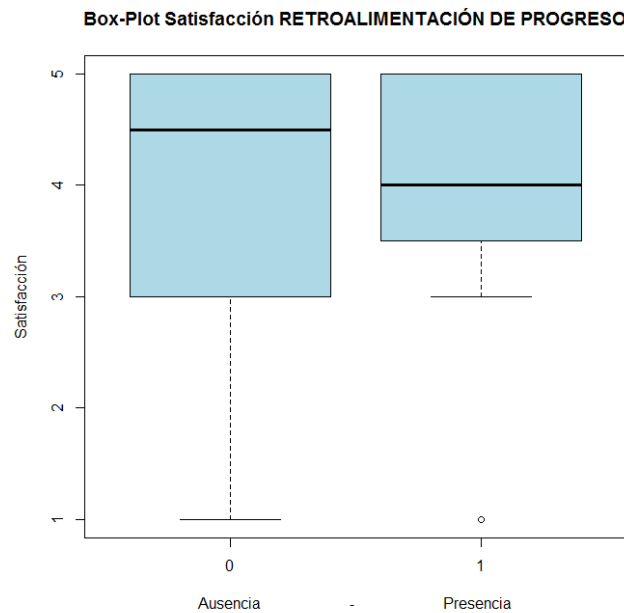


Figura 19: Diagrama de cajas para la variable Satisfacción en el Mecanismo de Usabilidad Retroalimentación de Progreso

C.3 Diagrama de cajas del Mecanismo de Usabilidad Preferencias

Observando la Figura 20 (Anexo C) podemos ver cómo la caja correspondiente a la ausencia del mecanismo de usabilidad tiende más hacia arriba, lo cual nos muestra que los sujetos con el mecanismo ausente tienden a dar más clics que los sujetos con el mecanismo presente. Notamos además la presencia de datos atípicos en ambos casos, tres en el caso de ausencia y cinco en el caso de presencia.

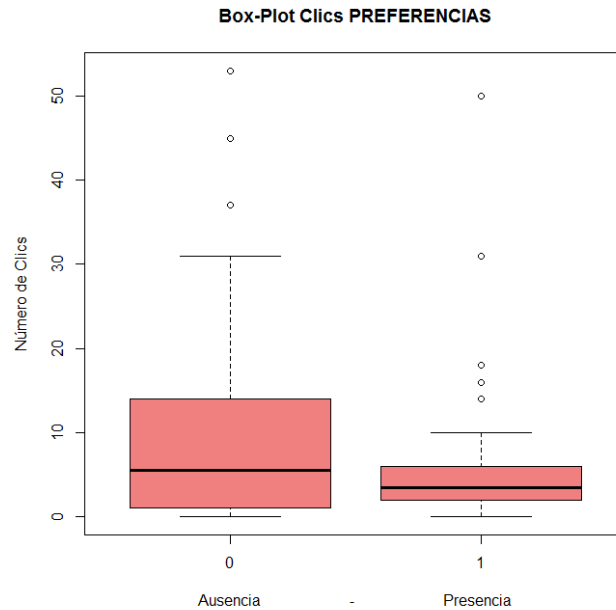


Figura 20: Diagrama de cajas para la variable N° de Clics en el Mecanismo de Usabilidad Preferencias

La Figura 21 nos muestra unas cajas muy similares a las correspondientes al número de clics, dónde el bigote correspondiente al máximo en el caso de ausencia del mecanismo llega a valores superiores. Esto nos da a entender que los sujetos que no poseían el mecanismo parece que utilizaban más tiempo a la hora de realizar la tarea. El diagrama correspondiente a la ausencia posee tres datos atípicos mientras que el correspondiente a la presencia tiene dos.

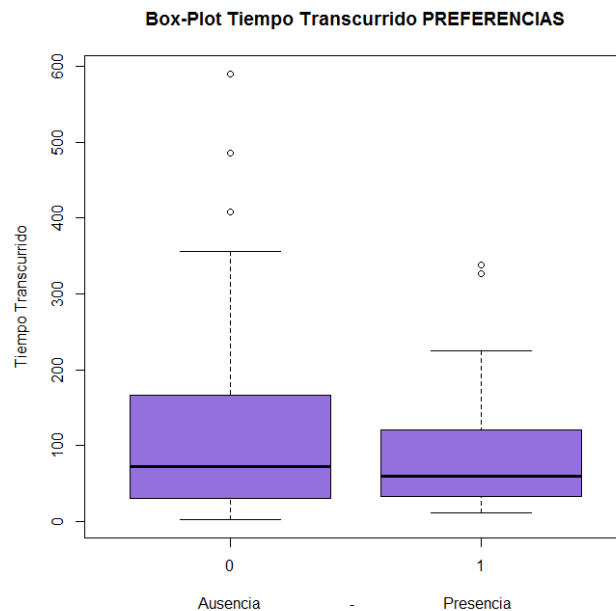


Figura 21: Diagrama de cajas para la variable Tiempo en el Mecanismo de Usabilidad Preferencias

Observando el *boxplot* de la Figura 22, podemos ver como las cajas son iguales pero invertidas, carecen de bigotes y la línea que corresponde a la mediana en el caso de la ausencia está en el cero mientras que la de la presencia está en el cien. Esto nos muestra que los valores varían entre cero y cien constantemente.

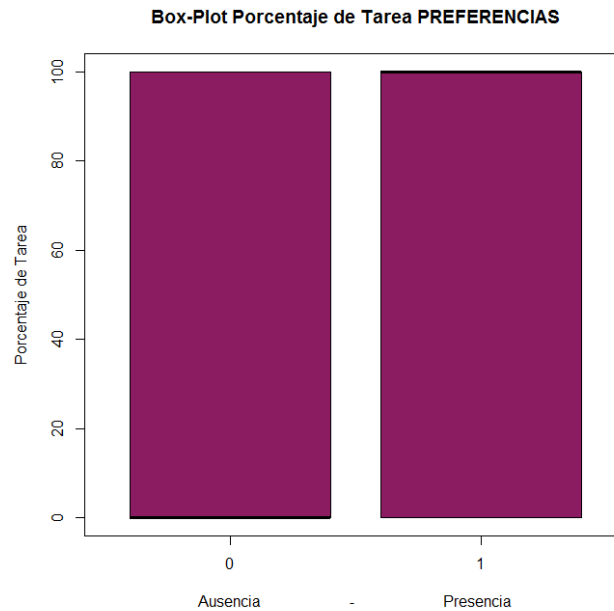


Figura 22: Diagrama de cajas para la variable Porcentaje de Tarea en el Mecanismo de Usabilidad Preferencias

En el diagrama de cajas de la Figura 23 se aprecian que las muestras difieren bastante y no reflejan simetría alguna debido a que el tercer cuartil de la primera muestra está casi alineada con el primer cuartil de la otra. También, la gráfica del grupo que ha tenido el mecanismo presente refleja un sesgo a la izquierda sin bigote superior, es decir, el tercer cuartil coincide con el valor máximo (ver Figura 23). En este caso se observa únicamente un valor atípico correspondiente a la muestra que ha tenido el mecanismo ausente.

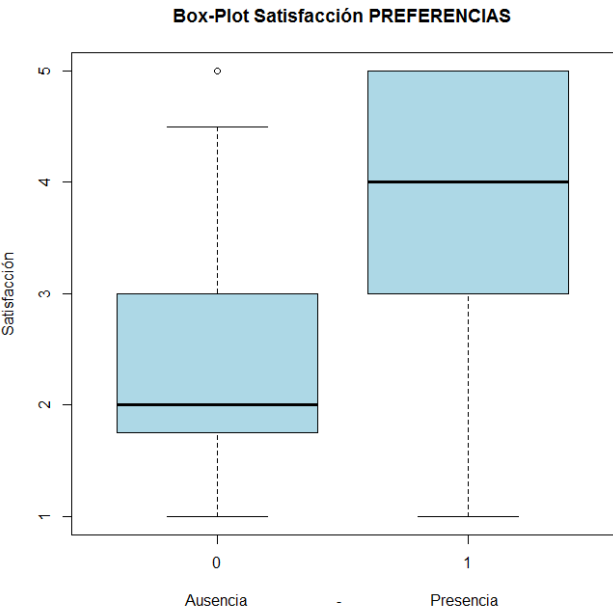


Figura 23: Diagrama de cajas para la variable Satisfacción en el Mecanismo de Usabilidad Preferencias